Spring 1-1-2011

Tuning and Control of Human Locomotion

Kristine Lynne Snyder
University of Colorado at Boulder, kristine.snyder@colorado.edu

Tuning and Control of Human Locomotion

by

Kristine L. Snyder

B.A., Bryn Mawr College, 2002

M.S., Integrative Physiology, University of Colorado, 2006

M.S., Applied Mathematics, University of Colorado, 2010

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Applied Mathematics

2011

This thesis entitled:
Tuning and Control of Human Locomotion
written by Kristine L. Snyder
has been approved for the Department of Applied Mathematics

_____

Rodger Kram

_____

Prof. James Meiss

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #0606.29

Snyder, Kristine L. (Ph.D., Applied Mathematics)

Tuning and Control of Human Locomotion

Thesis directed by Prof. Rodger Kram

Mathematical modeling and analysis have been an integral part of legged locomotion research for many years. While models from the very simple inverted pendulum model of walking and the spring-mass model of running to multi-segmental models with numerous muscles across each joint have been used to explore the process of legged locomotion, they are not always sufficient to explain how human locomotion adapts to a changing environment. Using techniques from dynamical and control systems, I : 1) identify the time scales involved in metabolic minimization in running, 2) explore how stability differs between walking and running, 3) develop an algorithm for optimal control in discrete physical systems, and 4) examine the changes in leg mechanics involved in uphill and downhill running. For the first project, I use ideas from control systems to identify the processes involved in metabolic minimization in running. For the second project, I use ideas of orbital and local stability, measured using Floquet multipliers and finite time Lyapunov exponents, to try to quantify dynamic stability in walking and running at preferred and transition speeds. For the third project, I define a new method for the constrained optimization problem underlying Discrete Mechanics in Optimal Control. For the fourth project, I use a control systems approach to examine the changes in leg dynamics from level to uphill and downhill running. By exploring the adaptations that occur with changing environment, I hope to reveal the mechanisms, and possibly some of the strategies, that lead to stable locomotion.

Dedication

To my husband, Steve, and to my late grandfather, Julian Herbert Hopkins, who gave up many of his own dreams so that I could have mine.

Acknowledgements

I would like to thank my advisors and current and former committee members, Rodger Kram, Jim Meiss, David Bortz, Alaa Ahmed, Juan Restrepo, and Jim Curry. I would like to thank my additional advisors Todd Murphey and Max Donelan for their sound advice and guidance, despite the fact I did not always make it easy for them. Further thanks go to my collaborators, Jinger Gottschall, Mark Snaterse, Chris Arellano, and Jason Franz. I would additionally like to thank my undergraduate assistants Timothy Dunn and Ignas Satkauskas and the Locomotion Labs at both University of Colorado and Simon Fraser University. Thank you to my husband, Steve, my family, and my friends for their support and my students for their inspiration and encouragement over the last 5-7 years. I could not have done it without you.

This research was partially funded by an NSF MCTP grant to the University of Colorado Applied Mathematics Department.

Contents

Chapter

Figures

Figure

Chapter 1

Introduction and Motivation

Because movement is such an inherent part of human life, there has been much interest and exploration of human locomotion using mathematical modeling. In the 20th century, walking and running were modeled as distinct gaits, each with its own model involving a specific energy-saving mechanism.

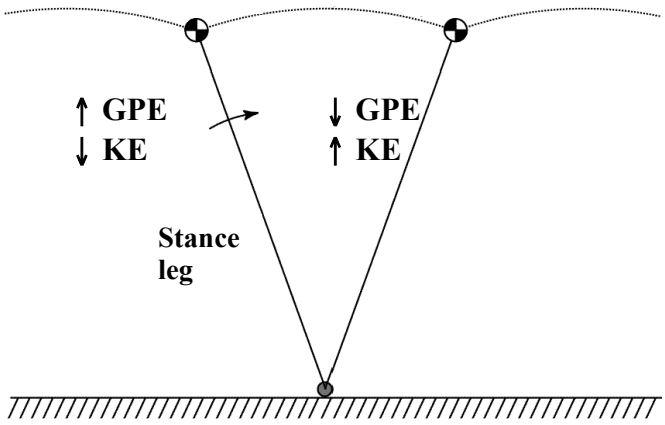

Figure 1.1: The inverted pendulum model

The classic walking framework is known as the "inverted pendulum," and is based on the assumption that during the single stance period (when only one foot is on the ground) the body vaults over the stationary foot, with the leg remaining rigid (Cavagna et al., 1963). At the beginning of stance, the center of mass (CoM) starts out at a relatively fast speed and thus has a large amount of kinetic energy, but the CoM height is relatively low and thus has a

modest amount of gravitational potential energy. Until midstance, when the CoM is directly vertically over the foot, as the body vaults upward and forward, kinetic energy is converted into gravitational potential energy. This gravitational potential energy is then reconverted into kinetic energy between midstance and the end of stance. Thus, the body can convert available kinetic energy to raise the CoM rather than having to expend metabolic energy to use muscles to do so. This mechanical energy is then recovered during the second half of stance. In reality, the muscles must still perform some additional work, but large energy savings are possible via the inverted pendulum exchange.

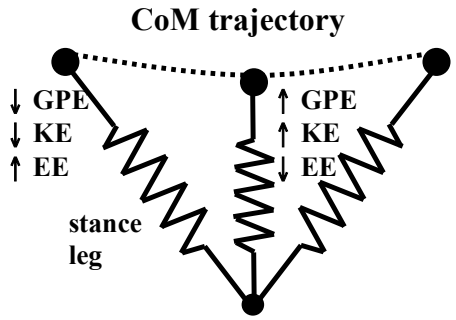

Figure 1.2: The spring-mass mechanical model

The classic running framework is known as the spring-mass model and is based on on a different energy saving mechanism. At the beginning of stance, the CoM starts out at a relatively fast speed, and thus has a large amount of kinetic energy, and the CoM height is relatively high and thus has a large amount of gravitational potential energy. From the beginning of stance to midstance, the CoM slows and lowers, as the hip, knee, and ankle joints flex, allowing the conversion of kinetic and gravitational potential energy into elastic energy, stored in the Achilles tendon and other elastic tissues. This elastic energy is then released between midstance and toe-off, as the ankle extends and the CoM moves upward and accelerates forward (Cavagna et al., 1977). As in walking, the muscles must still perform some work in running, but large energy savings are possible via energy conversion.

To understand how this elastic energy conversion can occur requires some knowledge of anatomy. Muscles produce force, which is transmitted by elastic tendons, which then pull on bone to rotate a joint. However, muscles do not have to shorten when producing force, defined as a concentric action. Muscle can also generate force while being lengthened by a greater external force (eccentric action), or generate force while remaining the same length (isometric action). In running, during the first half of stance (A to B), the gastrocnemius and soleus (calf muscles) produce force near-isometrically, i.e. while staying at the same length (Roberts et al., 1997; Lichtwark et al., 2007). The ground reaction force, the force applied by the ground to the foot, creates a torque on the foot that flexes the ankle joint and stretches the Achilles tendon, allowing the elastic energy storage to occur. During the second half of stance (B to C), the ankle joint extends and the tendon releases the stored energy, obviating the muscle work that might otherwise need to be performed.

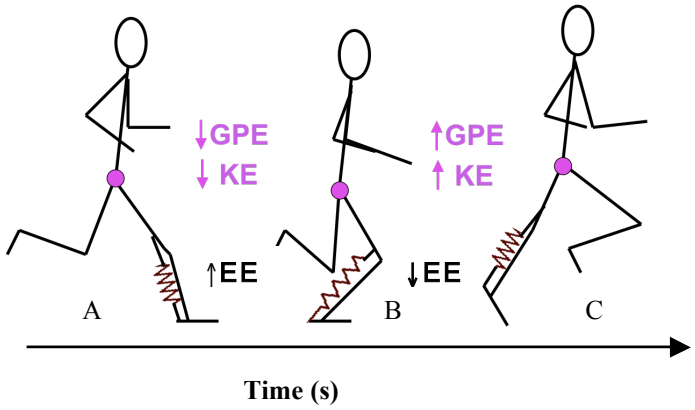

Figure 1.3: Relating human mechanics to the spring-mass model

The spring-mass model was almost simultaneously introduced by two different research groups. McMahon and Cheng (1990) introduced a system of ordinary differential equations that describes the spring-mass mechanical model and can effectively simulate running dynamics, both in terms of ground reaction force and the trajectory of the CoM. This model was further examined with respect to both mechanical and physiological constraints by Blick-

han (1989), including closed-form solutions for hopping. Later, an approximate closed form solution for running was derived by Geyer et al. (2005).

The spring-mass model has been sufficient and productive for studying level running and hopping under a number of different conditions. Variable stride frequencies (Farley and Gonzalez, 1996) were modeled by adjusting the overall stiffness of the leg, with a stiffer leg corresponding to a faster stride frequency. Changing surface stiffness (Ferris et al., 1999, 1998; Kerdok et al., 2002; Farley et al., 1998) showed that humans adjust their leg stiffness to keep the series stiffness of the surface and the leg constant. Deconstructing leg stiffness by joint into hip, ankle, and knee joint stiffnesses (Farley and Morgenroth, 1999; Arampatzis et al., 1999; Günther and Blickhan, 2002; Stefanyshyn and Nigg, 1998) revealed that stiffness is adjustable depending on the task at hand, with most adaptation occurring at the ankle joint.

Different approaches have been used to better understand the spring-mass model, each leading to robotic applications. A dynamical systems based approach, including looking at stability and bifurcation, was applied to this and other models by Holmes et al. (2006) and Ghigliazza et al. (2005). These researchers, along with Full, went on to apply these ideas to cockroaches, eventually successfully building and studying RHex, a series of hexapedal robots (Altendorfer et al., 2001, 2004; Koditschek et al., 2004)

The spring-mass model has also been explored thoroughly with respect to self-stability both passively (Seyfarth et al., 2002) and with control of swing-leg retraction (Seyfarth et al., 2003; Knuesel et al., 2005) and proprioceptive feedback (Geyer, 2003). These studies were reviewed by Blickhan et al. (2007). Seyfarth, Geyer, Blickhan and colleagues at the University of Jena went on to apply these ideas of feedback control to robotics in an attempt to build bipedal robots that consume less mechanical power (Iida et al., 2008; Seyfarth et al., 2009, 2006; Rummel et al., 2006; Iida et al., 2007; Rummel and Seyfarth, 2008; Rummel et al., 2008).

The inverted pendulum model has been similarly explored and extended based on ideas

introduced by McGeer (1990b), with researchers using a variety of mathematical models, human data, and robots to explore energy saving mechanisms in walking. The passive dynamic walking model started out as a simple inverted pendulum, with fixed rigid legs (Garcia et al., 1998). It was then adjusted to include modifications, such as knees or hips. Starting with very elementary models may sound futile because a simple model cannot encompass all of the details the movements required in human walking. But simple models offer some important insights because it allows researchers to compare modeling and experimental data. Because we cannot measure what happens internally, more complicated models are often not fully testable. Dynamic walking research integrates ideas from robotics, biology, and mathematics to explore control and stability in gait. Some of these models are purely mathematical, such as those done by Ruina and colleagues (Srinivasan and Ruina, 2005, 2007), and offer insight into what possibilities exist for stable gait. Some researchers, particularly Kuo's group at the University of Michigan, use these ideas to understand human gait, using simplified models and experimental work to look at how the body may accomplish energy savings (Kuo, 2001, 2002) and stability (Kuo, 1999). Still others, particularly Wisse's group out of the University of Twente in the Netherlands, have used the ideas from passive dynamic walking to design more effective robots, building more complex robots based on the energy savings mechanisms that seem to work in human walking (Collins and Ruina, 2006). Some modifications that have been made to the model have been to add knees (Collins et al., 2001), an upper body (Wisse et al., 2004), and adding leg swing (Wisse et al., 2005), either alone or together (Collins et al., 2005).

Other researchers, such as Mombaur of Heidelberg and Laumond of Toulouse have used more complex models, assuming the body is made up of rigid segments, to study human and robot locomotion from the optimal control perspective (Mombaur et al., 2010). These models do not always involve the drastic simplifications involved in the passive dynamic walking models, though they can. Rather, they usually include numerous anatomical joints, which are connected via actuation and spring-damper elements These models generally use

an inverse optimal control method in which the optimal control problem is solved via a bilevel technique (Mombaur et al., 2010). This methodology is unique because it does not insist on an *a priori* cost function. Note that here "cost" does not refer to metabolic cost, but generally some weighting of joint torques or actuation input values. Instead, the first level of this method describes the cost function as a weighted sum of distinct elements and solves for appropriate weightings of these elements by comparing them to measured values via least squares. In the second level, the appropriate state space and control trajectories are then solved for the cost function based on the previously found weights, typically via multiple shooting. The solution to this problem is then re-entered into the least squares problem to begin the next iteration. This methodology has proven effective for modeling human-like locomotion (Luksch et al., 2007; Mombaur, 2009). These models are open loop; they do not require any feedback of the state of the entire system to remain stable (Mombaur, 2007). However, only the second level problem involves derivatives, and it could be sped up significantly with an algorithm designed specifically for the appropriate control problem. Furthermore, the actuators used in this model do not take into account the relationship between length, velocity and force production that occurs in muscles.

Still other groups have examined what occurs on the muscle level, and used this data in order use forward dynamics to model the whole system of the lower or upper body (Pandy and Andriacchi, 2010). Delp's group out of Stanford has developed very popular open source software, known as OpenSim (Delp et al., 2007). In OpenSim, each of the muscles are modeled based on their specific isometric force-length and force-velocity relationships using a Hill-type muscle framework. Hill-type muscle models consist of a contractile element in parallel with a nonlinear spring, with this entire complex being in series with another nonlinear spring to represent each muscle (Delp and Loan, 2002). The algorithm finds joint forces based on matching experimentally identified trajectories (joint angles and angular velocities) and ground reaction forces with a least squares criteria. Because the kinematic and kinetic data are never entirely consistent, some correction is done on the mass parameters

and joint locations to allow for consistency. The algorithm then splits the forces between the synergistic muscles based on static optimization and drives a forward dynamic simulation via proportional-derivative control. The lower body model includes 43 muscle-tendon units. The upper extremity model includes 50 muscle-tendon units. These models are advantageous because they can be adjusted to individual subjects or pathologies. These models have been quite effective at modeling pathological gait due to stroke (Higginson et al., 2006) and cerebral palsy (Arnold et al., 2000, 2007; Hicks et al., 2007) because they allow the specificity necessary to examine pathology. There is also some research on the fundamentals of gait via these models (Fox and Delp, 2010; Liu et al., 2008). These examples give an idea of this method's utility, and numerous other projects have utilized OpenSim since it became available. However, OpenSim's methodology can be problematic because it represents an overdetermined system. That is, there are many different ways the muscle forces could combine in order to produce the output seen at the joints, and we cannot be certain the solution given by the model is the one the body actually uses. The actuation outputs of the model are often compared to empirical EMG (a measure of electrical activity in the muscle used to estimate muscle activation), which has its own measurement problems, being the sum of of individual sinusoidal action potentials. There is generally good agreement between the two, though timing and magnitude are never exactly comparable to EMG for all muscles. Because these models are searching for behavior for the entire system, this suggests that these models, though comprehensive, are not always accurate.

Another group using a similar forward dynamics simulation technique is Neptune's group at Texas. Rather than looking at the entire body, this group concentrates on answering specific questions about gait, which do not always require all muscles to be modeled individually. Like the group at Stanford, they use Hill-type muscle models, and use these to represent major muscle groups and/or major muscles. They use experimentally measured EMG patterns to restrict the timing at which different muscle groups are active. They use a simulated annealing algorithm to optimize muscle (contractile element) actuation based on

how well the simulation matched experimentally identified kinematics and ground reaction forces. This type of modeling has been used to examine muscle function in the transition from walking to running (Sasaki and Neptune, 2006a,b), during changing walking speed (Neptune et al., 2008), and during changes in the need for propulsion and body support (McGowan et al., 2009), to name a few. More recently, the Neptune group has examined control of various tasks in walking (Neptune et al., 2009).

Though all of these models are useful in specific situations, the more complex models often have more than one solution consistent with experimental results, whereas the simple models are not always sufficient to answer specific questions. Next, I present three ideas that are not taken into account in the simple models and difficult to measure using the more complex models, as well as offering an algorithmic adjustment that accelerates the process of finding optimal control. These ideas include the process behind metabolic minimization, an analysis of how stability changes across speed and gait, and the overall adaptations necessary for steady state uphill and downhill running.

The often unstated assumption behind almost all mathematical models of gait is the idea, long accepted in the field of biomechanics and energetics, that the main objective is minimization of the mechanical work done by muscles and hence metabolic energy expenditure. The inverted pendulum model is based around the energy saving mechanism of conversion of kinetic energy to gravitational potential energy and back. The spring-mass model is based on the energy saving mechanism of conversion of kinetic energy and gravitational potential energy to elastic energy and back. The muscle-tendon models in OpenSim and inverse optimal control involve elastic elements (Delp and Loan, 2002; Luksch et al., 2007). However, we do not know how this metabolic minimization is accomplished or what mechanisms are being used in order to gauge metabolic cost at any given time. The timing of this minimization could be critical to correctly model control in human gait. In Chapter 2, we use step frequency as a correlate for metabolic cost and explore the time scales involved in metabolic minimization.

Another important objective in bipedal locomotion is stability, which we define to be the dynamic equivalent of what in static situations is called balance. Measuring stability could help us analyze when human falls occur and what causes them, which is important for the health of the elderly and people with gait disorders. The main problem with stability studies has been that there is no accepted definition of stability, nevermind an accepted measure thereof. People have used mathematical measures for the models, such as eigenvalues, but have continued to use variability for experimental research, though recently, people have begun to use ideas from nonlinear dynamics (Hurmuzlu et al., 1995; Hurmuzlu and Basdogan, 1994). As a result, it is important to analyze stability using objective measures, such as those from dynamical and control systems theory to determine at what point within a stride people might be most unstable, and what kind of control might be required to stabilize them. Though nonlinear dynamics based stability has been explored very thoroughly in walking (Hurmuzlu et al., 1995; Hurmuzlu and Basdogan, 1994; Dingwell and Cusumano, 2000; Dingwell et al., 2000, 2001), only one study has examined running (Jordan et al., 2008).

An additional challenge for the inverted pendulum and spring-mass models occurs in explaining the transition between walking and running. In general, people walk at slower speeds (<1.75 m/s), and run at faster speeds (>2.5 m/s). Around 2 m/s, however, there is a range of speeds at which people can either walk or run. People typically avoid locomoting at speeds near this transition. If they have to use these speeds, people generally have a discrete speed at which they strongly prefer to switch from a walk to a run. This speed is called the preferred transition speed. For many years, it was accepted dogma in biomechanics that walking and running are distinctly different gaits with inherently different mechanics. Then Hof et al. (1983) and Hof (1998) proposed that elastic return also played a significant role in walking, and Geyer et al. (2006) adapted the well-known running model to walking. Geyer, Seyfarth, and Blickhan found that the paradigm of compliance so commonly used for running could also be applied to walking to more accurately predict center of mass dynamics. Though the two gaits were unified under a single model, the aforementioned model is too simple to

represent accurately the kinematic differences between walking and running, and has thus been insufficient to elucidate the motivations of the walk-to-run transition. In Chapter 3, we use nonlinear dynamics measure to explore dynamic stability in walking and running at both transition and typical speeds.

One of the additional problems with modeling walking and running is, as with all optimization problems, computational complexity. One method that has been used to model walking and running is "Discrete Mechanics in Optimal Control" or DMOC (Junge et al., 2005; Marsden and West, 2003). This method uses ideas from Lagrangian and Hamiltonian dynamics to find the control and trajectory variables that minimize a cost functional for a discrete physical system (Pekarek and Marsden, 2008). Though useful for modeling systems involving impacts, such as running and walking, this method is only as effective as the algorithm used to solve the constrained optimization problem it produces. While sequential quadratic programming (SQP) has been used in the past, this black box algorithm ignores some of the characteristics of DMOC that could be used to speed up convergence. By taking the parallel structure of DMOC into account in the algorithm, the convergence could be improved significantly. In Chapter 4, we offer an alternate method of solving DMOC by creating a projection based on the underlying parallel structure which can be solved using unconstrained rather than constrained optimization.

The spring-mass model is derived from a hopping model and is adjusted to allow running by adding forward motion. Due to its hopping-based derivation, however, one of the main tenets of the model is that the mechanical energy stays constant: no energy is added, and none is dissipated. Because of the energy conservation assumption, the classic spring-mass model is insufficient to explain uphill running, during which energy must be added to the system, or steady speed downhill running, during which energy must be dissipated from the system. In order to accurately characterize the mechanics and metabolics of uphill and downhill running, it is essential to adjust the spring-mass model for uphill by allowing for energy input, and downhill, by allowing for energy dissipation. With the addition of

an actuation element, one could characterize the work input necessary to accommodate a change in slope. In Chapter 5, we add actuation elements to the spring mass model to allow for uphill and downhill running to explore the input required from the body to adjust for a change in slope.

In this dissertation, I use mathematical techniques to analyze experimental data and modify models to account for the metabolic minimization, stability, walk-to-run transition and incline/decline. Using the results from these studies, I hope to help elucidate how the body and the brain function together to make stable bipedal gait possible in a variable and unpredictable world.

Chapter 2

Distinct fast and slow processes contribute to the selection of preferred step frequency in running

Collaborators: Mark Snaterse and J. Maxwell Donelan, Simon Fraser University

2.1 Introduction

A fundamental principle underlying locomotion physiology is that people select gait patterns that minimize energetic cost (Alexander, 1989). For instance, humans and other animals choose the gait that minimizes metabolic energy expenditure for a given speed of locomotion (Holt et al., 1995; Mercier et al., 1994). Within both walking and running, people choose the step frequency that minimizes their energy use (Cavanagh and Williams, 1982; Gutmann et al., 2006; Hogberg, 1952; Holt et al., 1991, 1995; Kaneko et al., 1987; Minetti et al., 1995; Umberger and Martin, 2007; Zarrugh and Radcliffe, 1978). More generally, while people can certainly choose to walk or run in many different ways, with only one recorded exception (Hunter et al., 2010), people consistently choose gait variables that minimize metabolic cost. In this paper, we explore the processes that underlie metabolic energy minimization in human running.

Recent research on walking has suggested that there are at least two distinct processes that underlie metabolic energy minimization. Snaterse et al. (2011) perturbed walking subjects with rapid changes in treadmill speed and observed the time scales involved in the subsequent adjustments to their step frequency. The authors identified a fast response in

step frequency that occurred within the first few seconds of a change in treadmill speed (Snaterse et al., 2011). This fast process dominated the overall response, accounting for about two thirds of the total adjustments to step frequency. Importantly, the fast time scale is too rapid to involve any direct feedback of metabolic energy expenditure. A second set of experiments demonstrated that this fast process encoded the relationship between speed and frequency that minimized energetic cost and that people persistently adjusted step frequency towards this frequency for each walking speed (Snaterse et al., 2011). Taken together, these results suggest that people rapidly approximate the energetically optimal walking pattern using a pre-programmed response based on prior knowledge of the relationship between body state and the associated metabolic cost. For instance, it is possible that the feedback gains of spinal reflexes or the timing circuits of central pattern generators inscribe the energetically optimal gait parameters (Pearson, 2004). Given a change from preferred gait parameters, feedback from vision, proprioceptors or other sensory receptors could all possibly activate pre-programmed gaits.

Snaterse and colleagues found that walking subjects also exhibited slow process that gradually adjusted step frequency over the first 30 seconds and was responsible for about one third of the total change (Snaterse et al., 2011). The timing of this slow process is consistent with direct optimization of energy expenditure (Figure 2.1B-D). Here we define direct optimization as the minimization of energy expenditure based on feedback from physiological sensors that sense signals directly related to metabolic activity. Direct optimization is expected to be slow for at least three reasons. First, candidate direct sensors of metabolic cost, such as chemoreceptors located in the medulla oblongata and the carotid and aortic bodies (Bellville et al., 1979; Fatemian et al., 2003; Kandel et al., 2000), as well as Group IV muscle afferents (Kaufman and Hayes, 2002; Kaufman et al., 1983), require at least 5 seconds to produce physiological responses to a stimulus. Second, instantaneous measures of energetic cost are not particularly representative of steady-state expenditure, which is best assessed by averaging over at least one stride. Finally, the energy expenditure sensed at one

particular step frequency does not necessarily indicate which other frequency will ultimately be optimal. It may be necessary for the person to iteratively adjust their step frequency, in a process that only gradually converges to an optimum. While direct optimization may contribute to the selection of gait patterns that minimize metabolic cost, the compounded effects of delays, averaging, and iterative convergence result in a slow optimization process.

The purpose of our research was to test if these fast and slow processes represent control mechanisms specific to walking or whether they instead represent a more general strategy for minimizing energetic cost in human locomotion and may therefore also translate to running. Consequently, we perturbed running subjects to determine how many processes were present in running and whether they corresponded to those found in walking. We focused on measuring step frequency adjustments relative to the preferred frequency because the preferred frequency for any given speed generally minimizes metabolic cost, and there is a well-established energetic penalty for frequencies faster or slower than the preferred value (Cavanagh and Williams, 1982; Gutmann et al., 2006; Hogberg, 1952; Kaneko et al., 1987). In addition to using similar physical (forced response) perturbations to those used by (Snaterse et al., 2011), we also performed novel non-physical (free response) perturbations to determine if the fast process is just a physical response to treadmill acceleration. To accomplish this, we used a metronome to enforce step frequency at values above and below preferred and observed the dynamics that subjects exhibited in the return to preferred step frequency when the auditory stimulus was removed. To impose physical perturbations, we applied rapid changes in treadmill speed and examined the processes involved in the consequent adjustments to step frequency. We then compared the dynamics of step frequency adjustments resulting from the free response experiments to those resulting from the forced response experiments. We quantified the presence of the fast and slow processes, and identified whether the same processes were present for despite different perturbations. Finally, we compared the dynamics found in both our running experiments to those found in the previous walking experiments to determine whether these processes represent general control mechanisms for selecting

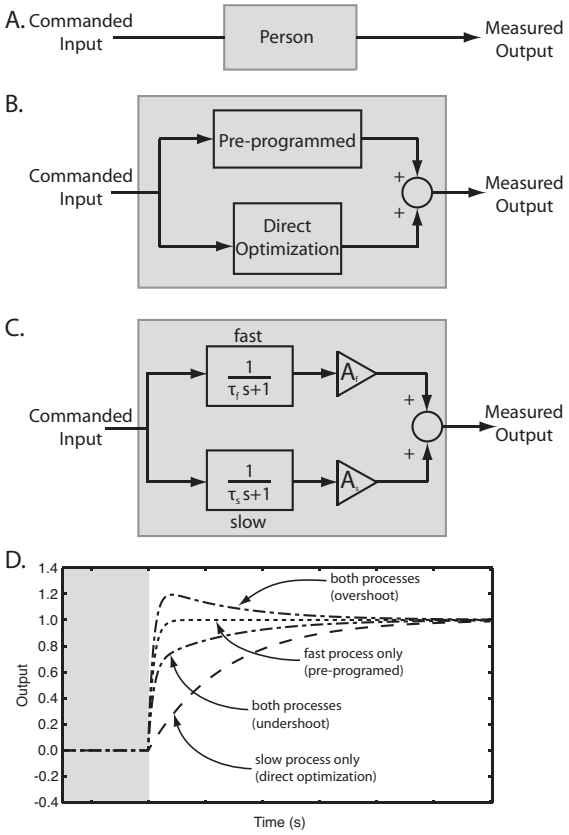

Figure 2.1: Dynamics of the System

A) We treat the person as a dynamic system that takes a commanded input and, through some process or combination of processes (grey box), translates it to a measurable output. B) We hypothesize that a combination of a fast process (pre-programmed) and a slow process (direct optimization) translate the input into the output. C) Mathematically, these processes can be represented by two transfer functions that act on two different time scales. D) We show the possible responses to a step input at time=0, which occurs at the end of the grey box. If only the fast process (pre-programming; dotted line) is active, the output rapidly reaches steady-state and never overshoots the steady-state value. If only the slow process (direct optimization; dashed line) is active, the output gradually approaches the steady-state value over the course of around 30 seconds. If both processes are active, there are two possibilities (dot-dash lines). The first possibility is an undershoot (dot-dash line), during which the output makes an immediate jump toward the steady-state value but does not quite reach steady-state, followed by a more gradual approach over a long time period until reaching the steady-state value. The second possibility is an overshoot (dot-dash line), during which the output makes an immediate jump past the steady-state value, followed by a gradual decrease to the steady-state value over a long time period. Whether an overshoot or undershoot occurs is determined entirely by the relative size of the amplitudes and not by the time constants.

energetically optimal gait patterns during human locomotion.

2.2 Methods

2.2.1 Subject and Equipment Information

Eleven subjects participated in this study. All subjects (6 female, 5 male; body mass 62.6 ± 9.2 kg; leg length 0.93 ± 0.05 m) were recreational athletes or members of the university track and field team. Simon Fraser University‘s Office of Research Ethics approved the protocol and all subjects gave written informed consent before participation.

Subjects ran on a treadmill (Trackmaster 425, Full Vision, Inc, USA) modified to allow the treadmill belt speed to be controlled by an analog input signal. The desired speed was dictated via computer in real time using a custom written program (Simulink Real-Time Workshop, Mathworks, Inc., Natick, MA). The actual speed was sampled at a frequency of 500 Hz using a magnet affixed to the treadmill flywheel and a reed sensor affixed to the treadmill chassis. We used pressure sensitive transducers fixed to subjects′ heels (for heel-strikers) or the balls of their feet (for mid-foot and forefoot strikers) and calculated step frequency from the time between consecutive foot-strikes. Footstrike characteristics were determined by asking subjects, and then verified during the warm-up.

Before any data were collected, we acclimated subjects using a 10-minute warm-up consisting of running on the treadmill at 2 m/s. Subjects were then asked to run briefly (30 seconds) at the fastest and slowest speeds required by the protocol detailed below to verify that they could sustain the full range of speeds and to allow them to become comfortable with speed changes.

2.2.2 Free Response Experiments

During the free response experiments, a constant step frequency was first enforced using the beat of an audio metronome, which was then replaced by white noise. Because

the treadmill speed was fixed, subjects were required to keep their average speed constant to remain on the treadmill. However, subjects were not required to change their step frequency once the enforced step frequency was released, and any change could occur over any time scale. We examined whether subjects adjusted step frequency when the enforced frequency was released, and identified the dynamics that contributed to any measured change.

Subjects ran at 3 m/s with step frequency first imposed by a metronome through headphones (Figure 2.2A). To allow us to test different magnitudes and directions of step frequency adjustments, we imposed four different step frequencies on each subject based on their preferred step frequency at 3 m/s. Two of these step frequencies were slower than the preferred value with the first equal to the preferred step frequency at 2 m/s and the second defined as twice the deviation from the preferred value at 3 m/s as the preferred value at 2 m/s. The other two step frequencies were faster than the preferred value with one equal to the preferred step frequency at 4.5 m/s and the other defined as twice the deviation from the preferred value at 3 m/s as the preferred value at 4.5 m/s. For instance, if a subject had a preferred step frequency of 2.8 Hz at 3 m/s and 3.2 Hz at 4.5 m/s, their highest imposed frequency would be 3.6 Hz. One subject was found not to vary step frequency between 2 m/s and 3 m/s, prompting us to remove him from further analysis. Each trial was 100 seconds in duration and each condition was repeated three times for a total of 12 trials. The time at which the metronome was switched off, and replaced with white noise, was randomly assigned to be 30, 40, or 45 seconds to prevent subjects from being able to predict when the metronome would stop. If a subject was unable to match the step frequency given by the last 10 seconds of stride frequency restriction, the trial was not used. However, this occurred for $<2\%$ of trials.

2.2.3 Forced Response Experiments

For the forced response experiments, we suddenly changed treadmill speed and examined its effect on step frequency. This perturbation required subjects to immediately change

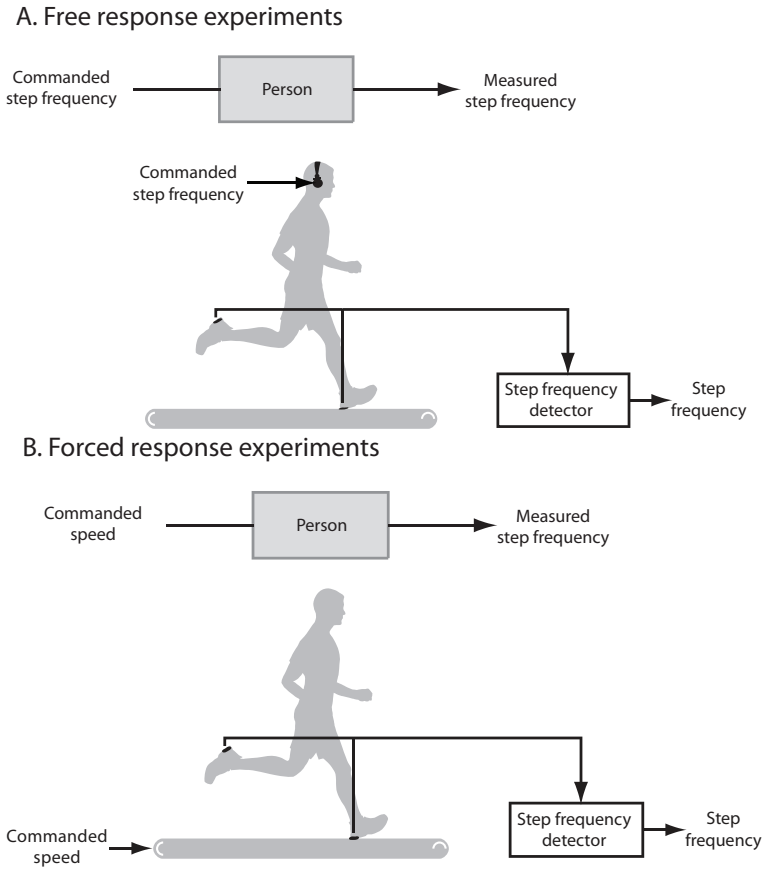

Figure 2.2: Experimental Methodology

A) In our free response experiment, subjects were restricted to a commanded step frequency with a metronome while treadmill speed remained constant. B) In our forced response experiment, we administered rapid speed changes to the treadmill, with subjects allowed to freely choose their step frequency. In both experiments, we measured the immediate and long term changes in step frequency that occurred as a result of the metronome turning off or the changes in speed.

speed in order to remain on the treadmill. However, they could change speed through any combination of step length and step frequency adjustments, including changing only step length keeping step frequency constant. Furthermore, these rapid physical perturbations did not require any slow step frequency adjustments. We compared the dynamics of step frequency adjustments in response to these physical perturbations with those identified in non-physical perturbations, with similar results indicating that common control mechanisms underlie the response to these distinct perturbations.

We imposed a series of step-like speed changes on subjects while they ran on the treadmill (Figure 2.2B). They began running for 90 seconds at 2.0 m/s and then were given 90 second periods of speeds at 2.5, 3.0, 3.5, 4.0, and 4.5 m/s in random order, each with a recovery period of 90 seconds at 2.0 m/s between the intervals. The maximum belt acceleration was set to 0.8 m/s^2 for both increases and decreases in speed. Depending on the magnitude of these perturbations, speed changes lasted for 0.8 to 5 seconds.

2.2.4 System Identification

We used standard techniques from system identification to quantify the dynamics of step frequency adjustments. System identification is a general term to describe algorithms for constructing mathematical models of dynamic systems from measured input-output data (Ljung, 1987). To prepare the data for this analysis, we used linear transformations to normalize each change in step frequency from 0 to 1 for each trial by subtracting the initial step frequency and dividing the result by the difference between the ending and initial stride frequencies. For instance, if the initial enforced step frequency was 3.0 Hz, and the ending step frequency was 2.8 Hz, these would convert to (3.0-3.0)/-0.2=0 and (2.8-3.0)/-0.2=1, respectively. This normalization allowed us to focus on the relative rather than absolute changes, and therefore infer the importance of each process.

To identify the parameters corresponding to the two processes, we used the free response experiments. We split the data from the free response experiments into two groups

based on their initial response to the perturbation. Preliminary analyses showed that some responses demonstrated an initial undershoot with the initial response approaching but not reaching the steady-state frequency (Figure 2.1D). Other trials demonstrated an initial overshoot, with the initial response overshooting the steady-state frequency. We used the first group, the undershoot data, to identify parameters corresponding to the dynamics of the two mechanisms, and the second group, the overshoot data, to test the generality of these parameters. Because the metronome was switched from a condition of being on to being off, we used a step function as the input for these free response experiments.

Based on previous research that had identified both a fast process and a slow process underlying the response to perturbations in walking Snaterse et al. (2011), we used a two process model for parameter identification. The mathematical representation of this two process model, expressed in the Laplace domain, takes the form

$$Y(s) = \left[\left(\frac{A_f}{\tau_f s + 1} + \frac{A_s}{\tau_s s + 1}\right) e^{-T_d s}\right] X(s), \tag{2.1}$$

where X(s) is the input and Y(s) is the output (Figure 2.1C). The parameters τ_f and A_f represent the time constant and amplitude for the fast process. Correspondingly, the parameters τ_s and A_s represent the slow process time constant and amplitude. The parameter T_d is a time delay to account for fixed physiological time delays, such as human reaction time. If the system input is an instantaneous step function of unit magnitude, and the system output is step frequency, f, the equivalent time domain expression is

$$\Delta f(t) = A_f(1 - e^{\frac{-(t-T_d)}{\tau_f}}) + A_s(1 - e^{\frac{-(t-T_d)}{\tau_s}}), \tag{2.2}$$

where t is time and the remaining parameters are as defined above. Figure 2.1 further illustrates how the output of this system in response to a step input is the sum of two exponential functions. The total response depends on the speed of the fast and slow processes as well as their relative contributions. We quantified the fast and slow processes using response time, defined to be the time required to achieve 95% of the total change for the

given process, as this measure better describes how long the process is active. The response time is approximately 3 times the time constant. We quantified the relative contributions of the two processes using the magnitude of the amplitude.

We first identified the parameters for the two process model using the measured step frequency response from the free response undershoot data. The identified parameters minimized the sum of the squared error between the model prediction and the measured step frequency adjustments for all undershoot trials. To calculate these best-fit parameters, we employed a gradient-descent based algorithm, seeded with an initial estimate of the parameter values. The identified parameters were insensitive to the initial estimates of parameter values. To implement this system identification, we used MATLAB′s built-in functions (e.g. ′idproc′ and ′pem′).

To test whether the measured adjustments to step frequency could be described by a simpler model, or if the dynamics were more complex than could be captured by our two process model, we also tested both one process and three process models. The degree to which the different models captured the measured responses was quantified by calculating R^2 values and by examining the residuals, defined as the difference over time between the model prediction and the measured data. We calculated R^2 values in two ways. The first calculation used the total error between the model prediction and the measured data for the individual trials by all subjects (individual fit). This is a rigorous restriction; in the two process model, for example, only five free parameters were used to describe the 43,086 total measurements from the free response undershoot data (10 subjects contributed 86 trials with each trial containing 501 data points). These comparisons led, perhaps deceivingly, to low R^2 values because the steady-state variability in step frequency was large relative to the step frequency changes induced by the perturbations. To reduce the effect of this steady-state variability on our goodness-of-fit metric, we also calculated the error between the model prediction and the average response across trials and subjects (average fit). This is still a strict test; the two process model used five free parameters to describe 501 data points

equating to 496 statistical degrees of freedom.

2.2.5 System Validation

To test whether the identified processes were used consistently across all free response trials, we determined how well the two process model predictions fit the measured overshoot data. We fixed the time constants and the time delay identified from the undershoot data, not allowing these parameters to vary and searched for the best-fit amplitude parameters. The two time constants were fixed because we wanted to test whether the same processes occurred. The time delay was fixed to more easily allow us to identify if there was an additional effect of the physical perturbation. We did not fix the two amplitudes because we had no *a priori* prediction concerning the relative contribution of the two processes. We assessed model fit by calculating the residuals and the R^2 values for both the individual and average data.

We also tested whether the processes identified from the free response trials explained the measured responses to rapid physical changes in treadmill speed. We first eliminated any forced response trial that had a step frequency change of smaller than 0.03 Hz, a value within the noise of the step frequency measurement. Because subjects‘ step frequencies do not always change very much at slow speeds, this did occasionally occur. Both the measured change in treadmill speed and the measured change in step frequency were then normalized to 1 for all trials as described earlier. Next, we binned the remaining data according to whether the initial response undershot or overshot the steady-state value. We then determined how well the two process model, identified from the free response undershoot trials, predicted the measured step frequency adjustments in response to this distinct perturbation. We fixed the time constants and the time delay parameters that were identified from the free response undershoot data, and searched for the best-fit amplitude parameters. The normalized treadmill speed was the input into this system identification. As with the earlier comparisons, we assessed model fit by calculating the residuals and the R^2 values for both

the individual and the average data.

2.3 Results

2.3.1 Free Response Experiments

When step frequency was enforced and then released, after a short delay (T_d=.37 ± 0.02 seconds, mean ± s.d.), subjects exhibited rapid changes in step frequency followed by longer-term adjustments that gradually brought step frequency to its steady-state value. This pattern in the undershoot data, where the initial adjustments in step frequency initially undershot the steady-state value, was well described by the two process model (Equation 2; Figure 2.3A). The identified response times associated with each process differed by more than an order of magnitude, with values of 1.47 ± 0.09 seconds (mean ± s.d.) for the fast process and 34.33 ± 0.87 seconds for the slow process. The fast process dominated the total response, the identified fast and slow process amplitudes were 0.67 ± 0.02 and 0.33 ± 0.02, respectively. The R^2 value for the average fit was 0.97, indicating that the model explained 97% of the average subject behavior. The R^2 value was lower for the individual fit, 0.36, because steady-state variability in step frequency was large relative to the step frequency changes induced by these non-physical perturbations. The residual errors also indicated that the two process model was a good fit, the errors were small in magnitude, randomly distributed around zero, and showed no particular pattern with time (Figure B).

Comparing the two process model fits with those from alternative models indicated that the simpler model was too simple, and a more complicated model was not needed to explain the measured results. A one process model was not sufficient to account for the observed adjustments in step frequency in response to the non-physical perturbation, leading to large residual errors that showed a distinct pattern over time (Figure 2.3B). This was also reflected in the R^2 values for individual fits, which decreased 45% from 0.37 for the two process model to 0.25 for the one process model. Similarly, the R^2 value when comparing

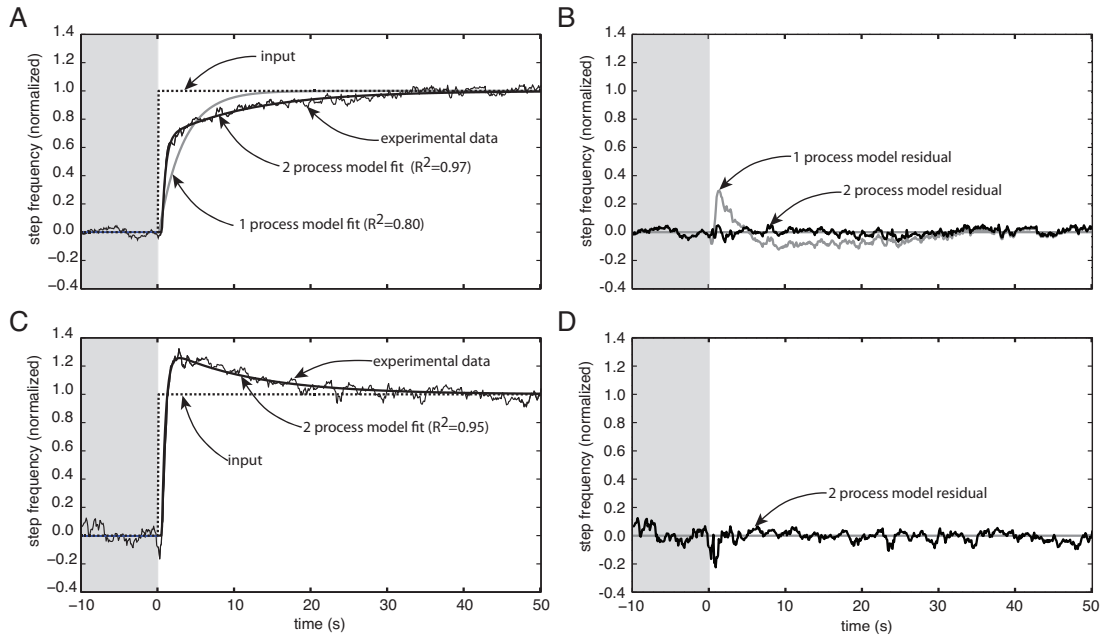

Figure 2.3: Free Response Results.

In all graphs, the grey rectangle is used to indicate when the metronome was on. Furthermore, though the fit was performed on all individual trials, the data shown (thin black line) in A) and C) is the average over all trials. A) 1 process (grey line) and 2 process (thick black line) models were fit to the undershoot experimental data using a step input (dotted line) to represent the change in the metronome signal. B) The residual error between the model and the experimental data shows that the 1 process model (grey line) was not sufficient to describe the dynamics. We also fit a 3 process model, but it was so similar to the 2 process model (black line) that it could not be shown without obscuring the 2 process data. C) The time scales identified for the undershoot data were fixed and the amplitudes varied to find a 2 process fit (thick black line) for the overshoot data (black line), again using a step input (dotted) line to represent the metronome. D) The residual error shows that this fit also very closely matched the experimental data.

average fits decreased from 0.97 for the two process model to 0.80 for the one process model. The more complicated three process model did not provide any additional information when compared to our two process modelthe R^2 values remained at 0.37 and 0.97 for the individual and average fits, respectively. Taken together, these comparisons suggest that a two process model is the simplest model required to describe the measured dynamics.

The two process model also accurately described the step frequency adjustments that initially overshot the steady-state value, indicating that the identified processes were used consistently across all free response trials (Figure 2.3C and D). This is evident from the small changes in R^2 values which decreased only slightly to 0.96 from 0.97 for the average fit comparisons and increased to 0.54 from 0.37 for the individual fit comparisons. The goodness of fit was also evident from the low magnitudes, random distribution, and lack of pattern observed in the residual errors (Figure 2.3D). The quality of this fit was particularly impressive given that the time constants and time delay parameters were fixed at the values identified from the undershoot data leaving only the two amplitude parameters to vary when fitting the overshoot data. For this overshoot data, the fast and slow process amplitudes were 1.33 $\pm$ 0.01 and -0.33 $\pm$ 0.01, respectively. However, note that because the ampltiudes are calculated from both fixed and varying parameters, these may be an understimate. Thus, in both undershoot and overshoot free response data, the fast process brought the step frequency within 33% of the steady-state value while the slow process fine-tuned the result.

2.3.2 Forced Response Experiments

Subjects exhibited similar behavior in the forced response experiments as in the free response experiments. There was a fast response followed by a longer-term adjustment of step frequency to its final value. The time constants identified from the free response data were, overall, a good fit to the data measured in this distinct experimental perturbation (Figure 2.3.2) with R^2 values of 0.67 and 0.87 for the average fits and 0.19 and 0.38 for undershoot and overshoot, respectively. As with the comparison to the free response overshoot data,

we made model predictions for the physical perturbation experiments by keeping the time constants and time delay parameters fixed at the values identified from the free response undershoot data, and leaving only the two amplitude parameters to vary. The identified amplitudes were similar between the two experiments, with undershoot values of 0.78 ± 0.01 and 0.23 ± 0.01 for the fast and slow processes, respectively, and overshoot values of 1.40 ± 0.01 and 0.40 ± 0.01 for the fast and slow processes.

The two process model identified from the free response experiments did not entirely explain the adjustments in step frequency in response to the physical perturbation, there were some additional dynamics that occurred within the first few seconds (Figure Figure 2.3.2B and D). This difference was not unexpected, while the metronome provided an impulsive auditory perturbation, the treadmill provided a physical perturbation that was stretched out over a finite period of time. These additional dynamics took place during the speed perturbation, and the residual errors paralleled the acceleration of the treadmill, indicating that they may simply reflect a biomechanical response to the treadmill acceleration (Figure 2.3.2B and D). These additional dynamics did not replace those resulting from the fast and slow processes, but supplemented them.

2.4 Discussion

Our results indicate that distinct fast and slow processes contribute to the selection of energetically optimal gait patterns during human locomotion. The fast process dominates the overall response to perturbations, completing 2/3 of the total step frequency change. The slow process takes about 20 times longer to fine tune step frequency and complete the return to the energetically optimal gait. This is a robust finding: we identified the same two processes irrespective of whether subjects overshot or undershot the steady-state value or whether the perturbation was non-physical or physical.

The processes that we identified in running match those found in the only previous study to look at this phenomenon, suggesting that common mechanisms underlie metabolic

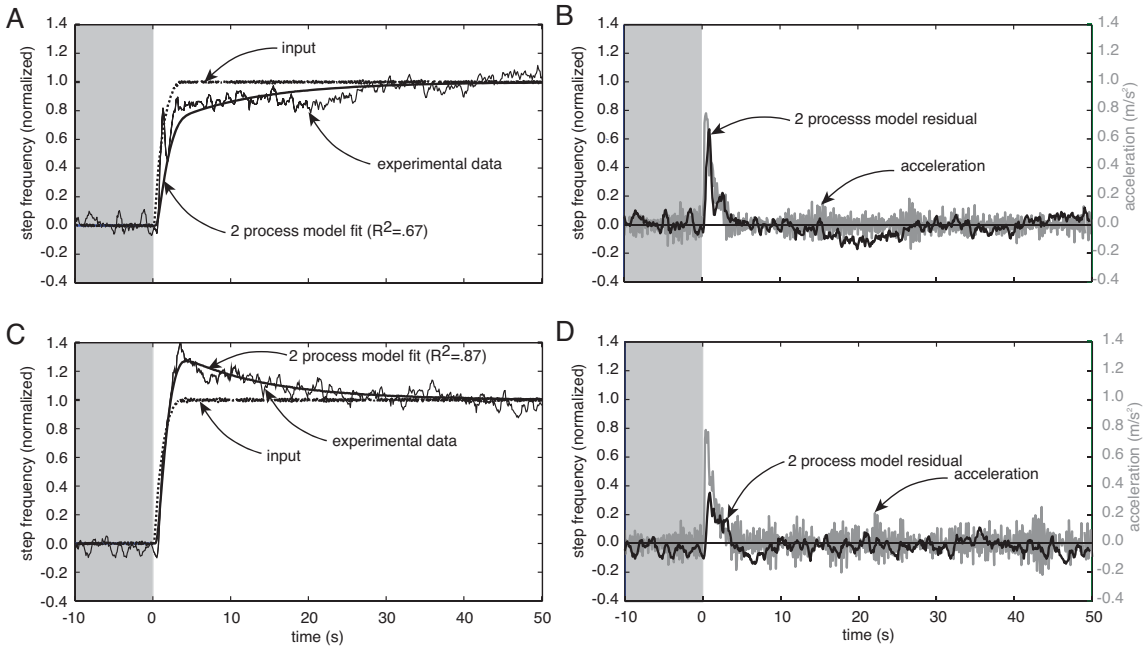

Figure 2.4: Forced Response Results

Though the fit was performed on all individual trials, the data shown (thin black line) in A) and C) is the average over all trials. In all graphs, the grey rectangle is used to indicate when the treadmill was at its initial speed. A) The 2 process fit using the time scales identified for the undershoot free response data (thick black line) also matched the forced response undershoot data well with the normalized treadmill speed used as input (dotted line). B) Residuals demonstrated that there were some additional dynamics (black line) involved in the forced response data that strongly paralleled the acceleration of the treadmill (grey line). C) The 2 process fit (thick black line) also closely approximated the overshoot forced response data. D) There were again some additional dynamics present (black line) that corresponded to the treadmill acceleration (grey line).

energy minimization across gaits. Snaterse et al. (2011) perturbed walking subjects by changing treadmill speed and identified a fast process response time value of 1.4 ± 1.1 seconds, which is very similar to runnings fast process response time of 1.5 ± 0.1 seconds. A similar correspondence is observed for the slow process response times (27.6 ± 16.2 seconds for walking and 34.3 ± 0.3 seconds for running). In addition, the relative contributions of the fast and slow processes were similar between walking and running. The fast process adjusted step frequency to within 34% of the final steady-state value during walking, and to within 23-40% of the final value during running, depending on type of perturbation. This suggests that not only do common neural mechanisms underlie metabolic energy minimization, but that the mechanisms are of comparable importance across gaits.

One difference between walking and running results was the distinct bifurcation in initial response to running perturbations, with some responses initially undershooting the steady-state step frequency while others initially overshot. While there was variability in the amount of overshoot or undershoot, these were clear categories, not arbitrary groupings of continuously varying variables. To further understand this pattern, we determined whether some subjects, or some conditions, were more likely to produce a response that overshot or undershot the steady-state value. The only general pattern that emerged was that subjects were far more likely to undershoot than to overshoot in both experiments ($p< 1.0 \times 10^{-4}$, t-test). 73% and 65% of the trials represented undershoots in the free response experiments and the forced response experiments, respectively. There were some additional experiment-specific effects. The direction of perturbation had a significant effect on the initial response in the free response experiments ($p< 1.0\times10^{-4}$), with subjects more likely to overshoot in drops in speed than in jumps. There were also statistically significant differences between subjects in their initial responses to physical perturbations (p=0.03), with some subjects tending to overshoot and some to undershoot. However, we did not find a direction-specific effect in the forced response experiments, or a subject-specific effect in the free response experiments, and neither experiment demonstrated a statistically significant effect of perturbation magnitude

on the initial response. While it is not clear why some subjects in some conditions initially overshot the steady-state value, the answer may be specific to running. In response to physical perturbations, walking subjects only initially undershot the steady-state frequency (Snaterse et al., 2011). Regardless of what type of behavior running subjects exhibited, the combination of short and long term processes still represented the observed dynamics very well.

There were a number of important limitations to our study. First, it is possible that our observations are specific to treadmill locomotion as the treadmill imposes a speed constraint that does not exist when moving overground. We have performed free response pilot experiments on subjects overground, and they suggest that our observations are not specific to treadmill locomotion. In addition, we have attempted to draw conclusions about metabolic energy minimization without directly measuring metabolic cost. This reflects a conscious decision to focus on collecting a wide range of perturbations (22 trials per subject). However, it also eliminated the possibility of having long duration trials that are required to accurately determine metabolic cost. Instead, we have relied on previous research by a number of different investigators, using a variety of experimental protocols, which have all demonstrated that the preferred steady-state step frequency minimizes metabolic cost (Cavanagh and Williams, 1982; Gutmann et al., 2006; Hogberg, 1952; Kaneko et al., 1987).

Our current experiments do not allow us to definitively conclude which physiological pathways are responsible for the fast and slow processes. Reflexes, central pattern generators, and descending commands from the brain may all play a role in both processes, and we cannot partition their contributions without further experiments. However, the results of our current experiments do exclude some important possibilities.

First, the processes are not simply biomechanical responses to a perturbation. This is most clear from the free response experiments where the perturbations were strictly auditory and all physical adjustments were self-induced. While there were physical perturbations in the forced response experiments, we observed additional fast adjustments to step frequency

that occurred during the physical perturbation, which supplemented, but did not replace, the fast and slow processes (Figure 2.3.2).

Second, the fast process we have identified is not the same phenomenon as the stumbling reaction reflex (Berger et al., 1984; Dietz et al., 1987). Previous studies employed conceptually similar treadmill belt speed perturbations to our physical perturbation experiments, but they used very fast accelerations that were more than ten times greater than those in our experiment (11.2 m/s^2 vs. 0.8 m/s^2). These accelerations were rapid and designed to challenge the balance of their walking subjects. Our accelerations were slow and designed to allow subjects to smoothly transition their speed. Our free response experiments clearly demonstrate that the fast process is not a balance correction response. We observed similar fast adjustments to step frequency even though balance was not challenged with a physical perturbation.

Finally, the fast process is too rapid to involve direct optimization of metabolic energy expenditure. The fast adjustments were essentially complete in under two seconds whereas feedback from physiological sensors that sense signals directly related to metabolic activity require at least five seconds to initiate physiological responses to a metabolic stimulus (Bellville et al., 1979; Fatemian et al., 2003; Kandel et al., 2000; Kaufman and Hayes, 2002; Kaufman et al., 1983). We consider this fast process a form of pre-programming because it contributes to producing the energetically optimal response without current knowledge of the actual energetic cost, relying instead on prior knowledge of the association between body state and metabolic cost (Figure 2.1B). This rapid response based on pre-programming may be accomplished by known mechanisms underlying the control of locomotion, including spinal reflexes and central pattern generators (Pearson, 2004). In contrast to the speed of the fast process, the approximately 30 second response time of the slow process is consistent with the expected timing of direct optimization of metabolic cost. As we described in the introduction, direct optimization is likely slowed by the compounded effects of feedback delays, averaging, and iterative convergence.

There are important energetic advantages to using both optimization and pre-programming in the control of step frequency. A key characteristic of optimization is accuracy, because it does not depend upon prior experience and will automatically adjust to novel circumstances, such as variable terrain or carrying a load, to converge on the energetically optimal gait. The accuracy of optimization provides one clear energetic advantage over using pre-programming alone. Inaccurate pre-programming can result in a steady-state difference between the selected and the energetically optimal gait whereas optimization will drive the steady-state error to zero. The magnitude of this energetic benefit will vary with the specifics of the situation as it depends on the precision of the pre-programming and how long the steady-state gait is maintained. The addition of a fast pre-programmed process also has a clear energetic advantage over using optimization alone in that it can better track the energetically optimal step frequency in response to continuously varying speeds. This advantage is largest for intermediate speed changes; optimization alone can track the optimal step frequency when speed is changing very slowly and neither process can adjust sufficiently fast when speed is changing very quickly.

To be more quantitative, we used our identified fast and slow processes to compare how a continuously varying speed affected the metabolic cost of running when using only the slow process to track the optimal step frequency for each instantaneous speed and when using both processes to track the optimal step frequency. In the model, we induced changes in step frequency by using a sinusoidal speed input to simulate the type of changes that might occur in steady state running. We used a sinusoidal input because research suggests than humans more likely evolved to run long distances than in short bursts. Furthermore, though a step input leads to significant energy savings in the short term, this savings decreases to zero within the first 30 seconds. We varied the amplitude of the sinusoidal input from 0.2 m/s to 2.0 m/s and the period from 0.1 to 20 seconds. We used a regression equation (Snyder and Farley, 2011) to calculate the percentage in change in net metabolic cost with a given percentage change in step frequency from preferred. Considering speeds that sinusoidally

oscillated between 2 m/s and 6 m/s, the difference between these two situations in their ability to track the optimal step frequency was maximized with sinusoid periods of 18 seconds. At this period, running using the slow process alone required a 4.4% increase in metabolic cost when compared to using both processes to select step frequency. The percentage difference is relatively small because the relationship between speed and energetically. Optimal step frequency is relatively flat in running. Even for large speed changes, the old optimal step frequency is not far from the new optimal step frequency (Cavanagh and Williams, 1982). However, the magnitude of this penalty is nontrivial; a 4.4% increase in metabolic cost when running at 4 m/s equates to roughly a 44 Watt penalty for a 70 kg runner. Furthermore, if conditions require variability at relatively fast speeds, any metabolic penalty may push the runner over their lactate threshold greatly reducing the duration of the run(Farrell et al., 1993) . Under the conditions in which running evolved, where humans may have often been at one side or the other of an extended chase (Bramble and Lieberman, 2004; Carrier et al., 1984), using a fast pre-programmed process to maximize the sustainable running speed may have been an important determinant of survival.

In summary, we found that there are two processes present in energy minimization in human locomotion. There is one short-term pre-programmed process that dominates the overall change and a longer-term process that serves to fine-tune this initial estimate. The same two processes are seen regardless of whether subjects overshoot or undershoot, and whether or not the perturbation is physical. Furthermore, the processes seen in running match the processes found in walking, both in timing and relative importance, suggesting that the mechanisms underlying these two processes are universal strategies for minimizing energy in locomotion. Moreover, including pre-programmed process along with the optimization process in response to variable speed can lead to significant metabolic savings. Future studies should explore the physiological mechanisms underlying these processes and the extent of their role in other, more novel situations.

Chapter 3

Analyzing Dynamical Stability in Walking and Running

Collaborators: Prof. Alaa Ahmed, Christopher J. Arellano, M.S., Ignas Sakauskas and Timothy Dunn

3.1 Introduction

Bipedal overground locomotion is categorized into two gaits: walking and running. These two gaits can be defined in two different ways. In the kinematic definition of walking, at least one foot is on the ground at all times. In running, for at least some portion of the step, neither foot is on the ground (aerial phase). In an alternate, biomechanical, definition, walking is defined as a gait in which the center of mass (CoM) is highest at midstance (when the hip is directly over the foot), whereas running is defined by the CoM being lowest at midstance (McMahon et al., 1987).

Previous modeling research by McGeer has shown that a passive walking model cycle recovers from small perturbations more quickly than a passive running model cycle (McGeer, 1990a,b). The difference in stability between the two gaits is indicated by smaller eigenvalues in walking associated with changes in speed, swing leg movement, and "totter", in which there is a mismatch between stride length and speed. Additionally, McGeer showed that in passive walking, increasing speed is associated with decreasing stability, whereas in running, increasing speed is associated with increasing stability in the form of decreased maximum eigenvalues (McGeer, 1990a,b). These models suggest that walking is more passively stable than running, and that in walking, increasing speed leads to a decrease in passive stability,

whereas in running, faster speeds lead to increased stability.

Though McGeer's research describes what happens quite well for a passive system, the human body is not passive. There is active control in human walking and running, but how much remains unknown. Certainly research done with decerebrate cats suggests the possibility that gait can be mostly controlled by a central pattern generator in the cerebellum or spinal cord (Whelan, 1996), though we do not have as much information on humans. If we can measure stability in walking and running, it is possible to make some inferences about control. For instance, in McGeer's passive models, walking is much more stable than running. If we were to find similar patterns of stability in actual human running, it indicates that similar control is involved in each gait. However, if we were to find running to be more stable than walking, it would indicate a greater influence of control in running than walking.

Because walking and running are mechanically different, many of the common biomechanical measurements cannot be compared across the two gaits, or at least no new information can be gleaned from such a comparison. For example, in walking, single support is generally considered to be the least stable time because having two feet on the ground for double support allows for more control. In running, single support is considered to be the more stable phase because the aerial phase offers no foot-ground interaction. However, by using more abstract mathematical ideas, it is possible to make some legitimate comparisons. One of these ideas is dynamic stability. Dynamic stability is a measure of how robust a system is in response to an outside perturbation. Because both gaits consist of cycles, nonlinear dynamics (NLD) has been used previously as a tool to analyze stability. NLD has been used extensively in walking (Hurmuzlu et al., 1995; Hurmuzlu and Basdogan, 1994; Dingwell and Cusumano, 2000; Dingwell et al., 2000, 2001), but, until recently (Jordan et al., 2008) this approach for examining stability had not been applied to running and has still not been studied at typical running speeds. By comparing measures of dynamical stability for walking and running for the same subjects, it would be possible to compare stability between the

two gaits. For experimental data we define stability to be the dynamic equivalent of what in static situations is called balance.

There are two different measures that can be used to examine stability with changing conditions of gait. The first measure is the Floquet multipliers of the overall state space of the system, commonly represented by lower limb joint angles and angular velocities. The Floquet multipliers measure how a perturbation from the limit cycle of the system changes as the system moves forward in time, and can be found by calculating the eigenvalues of the Jacobian of the function that represents how the system flows forward in time. It is only necessary to calculate the largest eigenvalue, as it determines whether the overall system is stable or unstable. A Floquet multiplier of magnitude greater than one indicates that a perturbation grows over time, indicating overall instability. A Floquet multiplier of magnitude less than one indicates that a perturbation shrinks over time, indicating overall stability.

The other measure commonly used to track dynamic stability in gait is the largest Lyapunov exponent (Dingwell and Cusumano, 2000). The Lyapunov exponents also track how perturbations change over time, but in a slightly different manner. The Lyapunov exponent tracks how two trajectories that start near to each other converge or diverge as they both move forward in time. Again, the largest value determines the overall stability of the system. Trajectories are generally assumed to diverge or converge exponentially, so a value greater than zero indicates instability, whereas a value smaller than zero indicates instability,

The essential difference between the quantities Floquet multipliers and Lyapunov exponents represent in gait studies is that Floquet multipliers represents the change between the stable trajectory and another trajectory. The Lyapunov exponent represents how two trajectories, neither of which is the stable trajectory, move toward or away from each other. It is possible for two trajectories to converge over time without converging to a stable limit cycle. However, if two trajectories converge to a stable limit cycle, they should ultimately

also converge to each other.

An additional way to compare stability and control across gait is to examine stability immediately after the transition from one gait to another. Though this has been done previously (Segers et al., 2006, 2007; De Smet et al., 2009), the only variables that were tracked were individual measures, such as stride frequency or center of mass dynamics. These studies did not examine measures of the overall system, such as eigenvalues of the joint angles of the lower limb. While individual factors such as stride frequency may give some information about the system, the overall dynamics would be more illuminated by measures that take the entire system into account. Additionally, these studies have all taken place while the subjects were undergoing acceleration, either on a treadmill or overground, which makes it impossible to determine what changes are due to speed as opposed to gait. Furthermore, these studies only involved fewer than 20 steps around transition, which may not be sufficient to show significant patterns.

Based on the results of the passive walking and running models, we hypothesized that we will find walking to be more stable than running. Furthermore, we hypothesized that lack of stability is one of the reasons that humans choose to avoid speeds near the preferred walk-run transition speed, and that therefore both walking and running will be more dynamically stable at typically preferred speeds than at the transition speed. To test these hypotheses, we used both Floquet multipliers and Lyapunov exponents as a way of measuring dynamic stability for each of these conditions, with lower values indicating greater stability for both variables.

3.2 Methods

3.2.1 Experimental Procedures

10 subjects (6 male, 4 female, mass = 66.9 $\pm$ 8.6 kg) participated in this study. Before any data collection, subjects completed a short questionnaire to determine whether they were

healthy and fit enough to complete the study. All subjects gave informed consent before participation in the study, and the protocol was approved by the University of Colorado Human Research Committee.

We used an 8-camera system (Motion Analysis, Santa Rosa, CA) to collect 3-D kinematics using 49 reflective markers at 250 Hz and a force plate (AMTI, Watertown, MA) instrumented treadmill to collect force data at 1000 Hz. After performing a static trial on the treadmill to determine the relative position of the reflective markers and subject mass, subjects completed 6 separate 6-minute trials in random order on a treadmill. The six trials were as follows: 1.25 m/s walking, 2 m/s walking, 2 m/s running, 3 m/s running and two additional trials at 2 m/s. 2 m/s is a typical walk-run transition speed (Hreljac, 1993). For one of these trials, subjects walked for 3 minutes at 2 m/s before switching to running for the remaining 3 minutes. For the other trial, the order of gaits was reversed.

The sacrum (commonly called the tailbone) marker was used as an approximate location of the center of mass, and was placed on the bony prominence on the midpoint of the bone, about 3-4 inches below the posterior superior iliac spines.

We calculated joint flexion and extension angles from the marker data. Because some of the reflective markers, particularly those at the hip, could be blocked by arm swing, we were sometimes required to rebuild the marker location using other visible markers, or, on rare occasions, using a spline approximation.

3.2.2 Eigenvalue Analysis

The kinematic data were then used to do a mathematical analysis of the dynamic stability of the system, using two distinct measures. The first is a Floquet analysis to measure the orbital stability of the dynamical system. We defined the system by the hip, knee, and ankle joint flexion-extension angles and angular velocities for the left leg. Assuming symmetry, left and right legs should give similar results. Orbital stability measures how quickly the system returns to its stable trajectory after an initial perturbation. The underlying assump-

tion for such a measurement is that a periodic limit cycle exists for the system. Though we cannot be certain that an exact limit cycle exists for walking, one approach is to normalize all trajectories to have length 101, and then average over all measured steps to get an approximate limit cycle (Dingwell and Kang, 2007). Under this assumption, the following relationship applies for any two subsequent state vectors i and $i+1$ for the Poincaré section at any given point of the stride:

$$x_{i+1} = F(x_i) \tag{3.1}$$

where, F is the map describing the model dynamics. Recall that, $x* = F(x*)$, where $x*$ is the equilibrium point, the point in this specific Poincaré section that is on the limit cycle. Using this identity, we can simplify by linearizing around the equilibrium point with the Jacobian

$$\delta x_{i+1} = J(\delta x_i). \tag{3.2}$$

The stability of the system is determined by the eigenvalues of the matrix J. If these eigenvalues are all of magnitude less than 1, any perturbation in the system will shrink with each step, indicating that the system is stable. If any eigenvalues of this matrix have magnitude greater than 1, then there is at least one direction in which the system is unstable. Thus, to check for overall stability, the largest magnitude of the eigenvalues must be less than 1.

Because the system we used is purely numerical rather than analytical, we constructed the Jacobian rather than calculating it. We used a previously defined method (Hurmuzlu and Basdogan, 1994) to do so for each of the 100 Poincaré sections. For each subject, for each point in the cycle, two matrices were formed made up of the vertical state space perturbation vectors (state-space vectors - stable trajectory), horizontally concatenated by step. The first matrix, A, was an $s-1 \times n$ matrix, where s is the number of steps taken, and n is the dimension of the state space, and represented steps 1 to s-1.. The second matrix, B was an $s-1 \times n$ matrix, which consists of the matrix A moved up by 1 row representing steps 2 to

s. That is row $i = x_i$ of B consisted of the elements of row $i + 1 = x_i$ of A,. The Jacobian was then calculated by using the least squares algorithm on the following equation for j=1 to s.

$$\delta x_{i+1}^{j} = \sum_{k=1}^{s} J_{j,k} \delta x_{i}^{j} \tag{3.3}$$

Finding the eigenvalues of J then determined the stability of the system at each percentage of a stride cycle. We then averaged over the entire cycle to quantify the stability over an entire stride. Because noise in the data can significantly effect the magnitude of the eigenvalues, we filtered the data at 20 Hz before performing this analysis.

3.2.3 Lyapunov Exponent

Additionally, we calculated finite time Lyapunov exponents for the sacrum and the left toe markers. The two values were found because the stability of the sacrum approximates that of the CoM and is thus most representative of the whole system. The toe was used because previous research has found that in walking, humans will change other leg kinematic patterns when perturbed in order to preserve toe clearance, indicating its importance in terms of stability (Ivanenko et al., 2002).

Because it is difficult to approximate Lyapunov exponents, which were originally infinite measures, by using finite data sets, calculating the Lyapunov exponents is mathematically very involved. We used a variation of the method put forth by Rosenstein et al. (1993) as outlined in the following paragraphs.

We first calculated the time delay by finding the first minimum of the average mutual information function (Fraser and Swinney, 1986). After determining the time delay T, we then used the false nearest neighbors analysis to determine the embedding dimension, m using the method outlined by (Kennel et al., 1992). Previous work has shown that though Takens requires an embedding dimension $< 2n$ (Takens, 1981), where n is the box-counting dimension, a smaller embedding dimension is often sufficient to sufficiently expand

the attractor (Rosenstein et al., 1993). The false nearest neighbors analysis determines the embedding dimension of a dynamical attractor based on the idea that, when the attractor is projected down into a dimension smaller than its actual dimension, some points will appear close together (i.e. will appear to be neighbors) that are not actually close together in the dimension of the attractor. These are considered false nearest neighbors.

To determine the actual embedding dimension, the percentage of false nearest neighbors is calculated in dimension d and then compared to the number of false nearest neighbors in dimension $d+1$. When the number of false nearest neighbors drops to within a certain tolerance (ϵ) of zero, that dimension is determined to be the embedding dimension of the attractor.

To estimate the number of false nearest neighbors, we calculated the difference between the Euclidean distance between 2 points in d and $d+1$ dimensions. If the dimension of the attractor really is d, then the distances should be the same, whereas if a neighbor is false, the distance will be larger in $d+1$ dimensions than in d dimensions. The square of the Euclidean d-norm of each point with its r^{th} nearest neighbor is determined by

$$R_d^2(n,r) = \sum_{k=0}^{d-1} [x(n+kT) - x_r(n+kT)]^2 \tag{3.4}$$

and the square of the Euclidean d+1-nrm is

$$\begin{aligned} R_{d+1}^2(n,r) &= \sum_{k=0}^{d} [x(n+kT) - x_r(n+kT)]^2 & (3.5) \\ &= R_d^2(n,r) + [x(n+dT) - x_r(n+dT)]^2 & (3.6) \end{aligned}$$

Subtracting these two and normalizing by the square of the Euclidean d-norm gives

$$\left(\frac{R_{d+1}^2(n,r) - R_d^2(n,r)}{R_d^2(n,r)} \right)^{\frac{1}{2}} = \frac{|x(n+Td) - x_r(n+TD)|}{R_d^2(n,r)} \tag{3.7}$$

If the value given by Equation (3.7) is less than the value given by R_{tol}, the neighbor is considered false. Kennel et al. (1992) give an R_{tol} value of 10.

Further exploration showed that another criterion is necessary for determining actual embedding dimension. Kennel et al. (1992) found that if noise was added to the system, it

led to a situation where a point was not actually that close to its nearest neighbor. That is $R_d(n) = R_d(n, r = 1) \approx R_A$ where R_A is approximately the diameter of the attractor. This is a natural result of having a finite number of points and expanding them into higher and higher dimensional space. Consequently, an additional criterion of

$$\frac{R_{d+1}(n)}{R_A} > A_{tol} \tag{3.8}$$

was included. Failure of any test was sufficient to determine that a neighbor is false.

We then did a state space reconstruction using the method of Rosenstein et al. (1993) on the original 1-dimensional data array $\{x_1, x_2, \ldots, x_N\}$, using the embedding dimension, m, and time delay T determined previously.

$$X_i(t) = (x_i, x_{i+T}, \ldots, x_{i+T*(m-1)}) \tag{3.9}$$

Thus the original 1 x N array was converted into an M x m matrix, where

$$M = N - (m-1) * T \tag{3.10}$$

After reconstructing the state space, nearest neighbors were determined using the following criteria

$$d_j(0) = \min_{X_j} \|X_j - X_{\hat{j}}\| \tag{3.11}$$

and

$$|j - \hat{j}| > mean\ period \tag{3.12}$$

The first criterion assures that $X_{\hat{j}}$ is the nearest neighbor in terms of Euclidean distance, whereas the second criterion determines that a nearest neighbor actually exists, in terms of there being a neighbor outside a period.

To allow for comparison across measures, we also calculated Lyapunov exponents for the state space used to calculate the eigenvalues, the joint angles and angular velocities. Because this representation of the system already includes six dimensions, we did not expand

the attractor using the method outlined previously, but performed the following analysis on this six dimensional data set.

In past studies at this point, the largest Lyapunov exponent λ_1 was then estimated. The definition of λ to approximate

$$d_j(i) \approx C_j e^{\lambda_1 (i\Delta t)} \tag{3.13}$$

where C_j is the initial distance between two nearest neighbor trajectories j and $\hat{j}$, i is the number of time steps elapsed, and Δt is the time step. Logging both sides gives

$$\ln d_j(i) \approx \ln C_j + \lambda_1 (i\Delta t) \tag{3.14}$$

Note that this is actually a set of M equations of i with slope λ_1, so we can find λ_1 by finding the average slope of these lines. We do so by using a least-squares method to estimate λ_1 from

$$y(i) = \ln C_j + \lambda_1 * i = \frac{1}{\Delta t} \langle \ln d_j(i) \rangle \tag{3.15}$$

where the average over all values is represented by $\langle\rangle$. Note that in actual experimental data, though the data may not be linear over the entire range of time, if this type of analysis is appropriate, there should be a clear range where the relationship is linear from which λ_1 can be determined. A value of $\lambda_1 < 0$ indicates stability.

However, for our data, we did not find a range of linearity after taking the logarithm. In fact, analysis of the original system without taking the logarithm revealed that, at least in this case, an exponential did not provide a good fit. We therefore performed some of our analysis on the raw data. Note, however, that the relative relationships in terms of stability would be the same regardless of whether we used the raw or logged data. We performed a sensitivity analysis on this data to determine whether or not the Lyapunov measure changed depending on the number of time steps used to calculate the value. Furthermore, because there was some noise in the output data for the first 100 points, we did not use these points for analysis.

Another measure of stability is how quickly two trajectories diverge from each other. We therefore examined the both the time and the number of strides to saturation. We defined time to saturation as the time it took for the slope of the divergence to decrease below a given tolerance level. Due to the noise in the data, we found the most consistent values when this tolerance was set to a value of 10^{-5}. Strides to saturation were found by finding the number of strides it took to reach this tolerance value.

Furthermore, initial results showed problems with the embedding dimension identified by the global false nearest neighbors algorithm. The algorithm was sometimes unable to find a dimension where the number of false nearest neighbors dipped below a reasonable tolerance level. While this is likely a sign of higher dimensional noise in the system, filtering the data showed further problems. This high dimensional noise was enough to mask an underlying change in the number of false nearest neighbors and consequent change in embedding dimension, as revealed by a comparison between filtered and unfiltered data. We therefore did a sensitivity analysis to identify the effect of different embedding dimensions on the Lyapunov exponent.

3.2.4 Stability Changes Across Transition

We did some additional preliminary analysis with the trials during which the subject switched from a walk to a run. Because this is a preliminary analysis, we have thus far only performed this anlysis on 6 subjects. We used the knee, hip, and ankle data to look at how stability changed immediately around the transition. To do this we used the last two minutes to define the stable attractor for the second gait as described previously. We then examined the behavior immediately after transition based on an eigenvalue analysis. To do this we used a moving window of 100 strides starting with the first 100 strides after the transition, and calculated the eigenvalues for heelstrike for each window as it stepped forward 1 stride for a total of 100 strides.

3.2.5 Statistics

We performed a 2 x 2 repeated measures ANOVA on the eigenvalues and Lyapunov exponents, with gait being defined as either walking or running and speed being typical or transition. We required $p<0.05$ for significance. We used Tukey's HSD to determine differences between different groups.

3.3 Results

3.3.1 Eigenvalue Analysis

We found significant differences between the average eigenvalues over the course of a stride across gait ($p = 2.7 \times 10^{-4}$; Figure 3.2), but found no significant effect of speed ($p = 0.1653$) or interaction ($p = 0.0801$). No other significant effects between conditions were found. To assure that the increase in stability from walking to running was not only due to the inclusion of the aerial phase, we also performed these analyses for stance, and for beginning of stance (heel-strike) and end of stance (toe-off). However, we found very similar patterns in overall eigenvalues for all conditions. We did find that the eigenvalues at heel-strike tended to be greater than the average values during stance, the values at toe-off, or the average values during the entire stride. Additionally, the eigenvalues for the faster speeds tended to be greater when averaged over a stride than for the other values.

3.3.2 Lyapunov Exponents

Joint Angles and Angular Velocities A Lyapunov analysis of the joint angles revealed similar patterns. The results showed a significant effect of gait ($p = 0.00013$), as well as a significant interaction effect ($p = 0.00071$). The Lyapunov measure values for walking at 1.25 m/s were significantly lower than those found for walking at 2 m/s ($p < 0.01$). The values for running at 2 m/s were significantly less than those for walking at 2 m/s ($p < 0.01$).

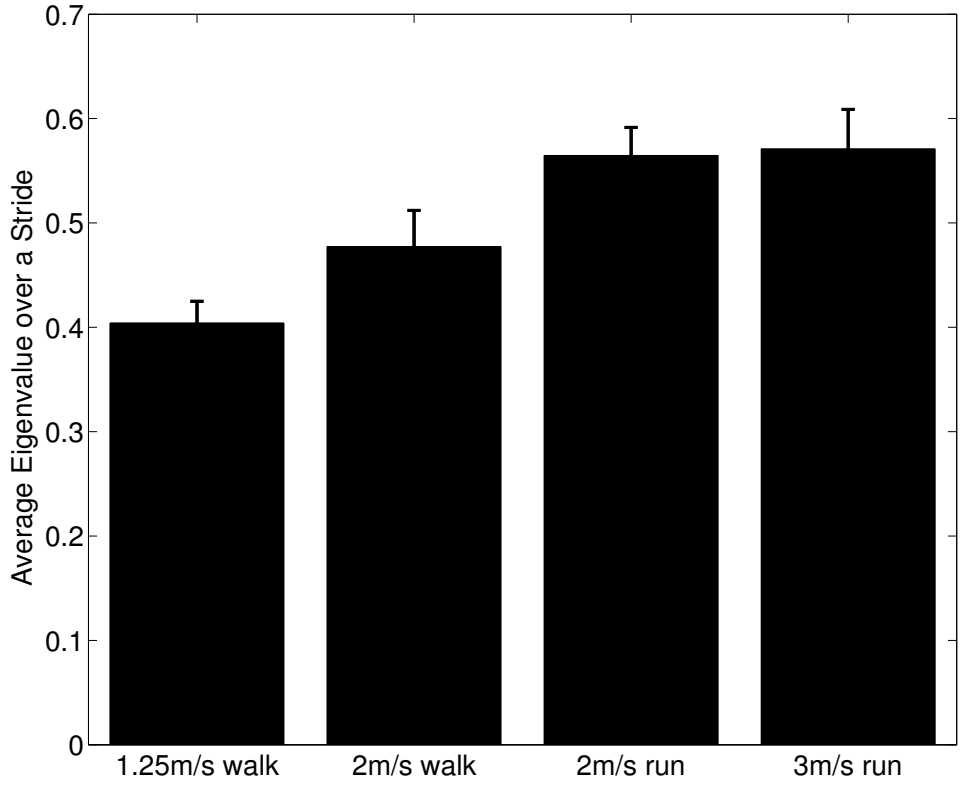

Figure 3.1: Average eigenvalues over the course of a stride.

The average eigenvalue over the course of a stride changed with condition, with an increase in eigenvalue being found between walking and running. Figure shows means ± SEM.

Additionally, the values for running at 3 m/s were significantly greater than those for running at 2 m/s ($p < 0.01$). Though these values come from the non-logged data for the first 10 strides, similar results were found whether we used strides 1-5, strides, 4-10, strides 1-10 to calculate these values. We also found similar results regardless of whether or not we log transformed the data to calculate the Lyapunov exponent.

Vertical Position of the Sacrum Analyzing the vertical position of the sacrum, we found a significant effect of gait ($p = 2.8 \times 10^{-6}$), as well as a significant interaction effect ($p = 0.0099$; Figure 3.4). The Lyapunov measure values for walking at 1.25 m/s were significantly less than those found in walking at 2 m/s ($p < 0.01$). Additionally, the values

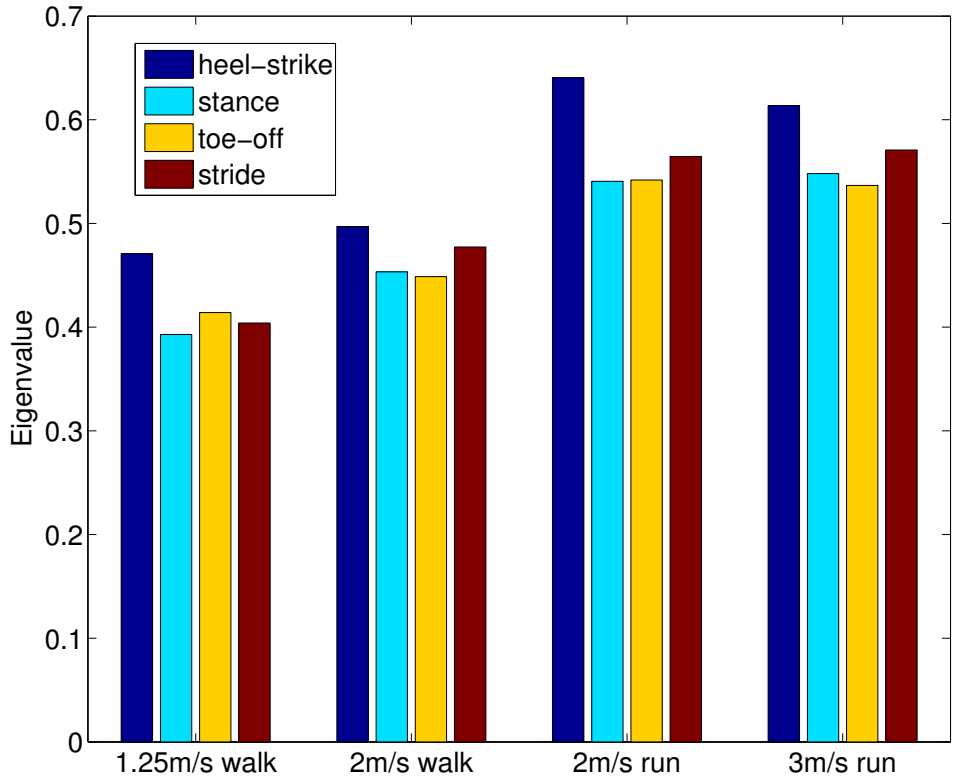

Figure 3.2: Average eigenvalues

We found a similar pattern of change with condition regardless of when during the stride the value is calculated. We also found higher values at heelstrike across conditions, with only small differences found between the overall average stride values and other times during the stride.

found in running at 2 m/s were greater than those found for walking at 2 m/s ($p < 0.01$). Furthermore, values for running at 3 m/s were significantly greater than those found for 2 m/s ($p < 0.01$). Though these values come from the non-logged data for the first 10 strides, similar results were found whether we used strides 1-5, strides, 4-10, strides 1-10 to calculate these values. We also found similar results regardless of whether or not we log transformed the data to calculate the Lyapunov exponent.

Mediolateral Position of the Sacrum Using the mediolateral (side-to-side) position of the sacrum, we found a significant effect of gait $p = 0.03$, as well as a significant

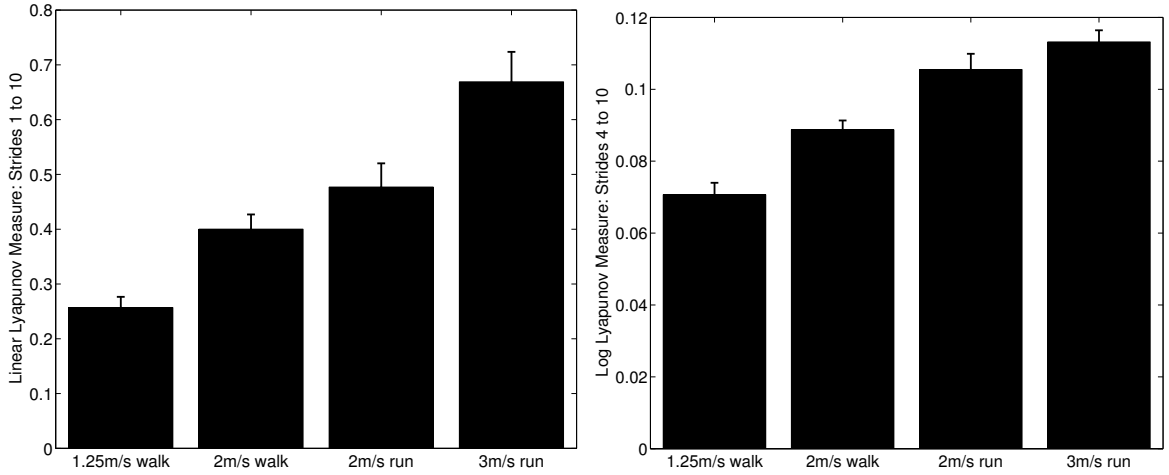

Figure 3.3: Lyapunov measures for joint angles and angular velocities.

The Lyapunov measure changed distinctly with condition. There was an increase from walking at 1.25m/s walking to 2m/s walking, and another increase from 2m/s walking to 2m/s running. Additionally, the value for 3 m/s running was greater than that for 2 m/s running. Figure shows means ± SEM.

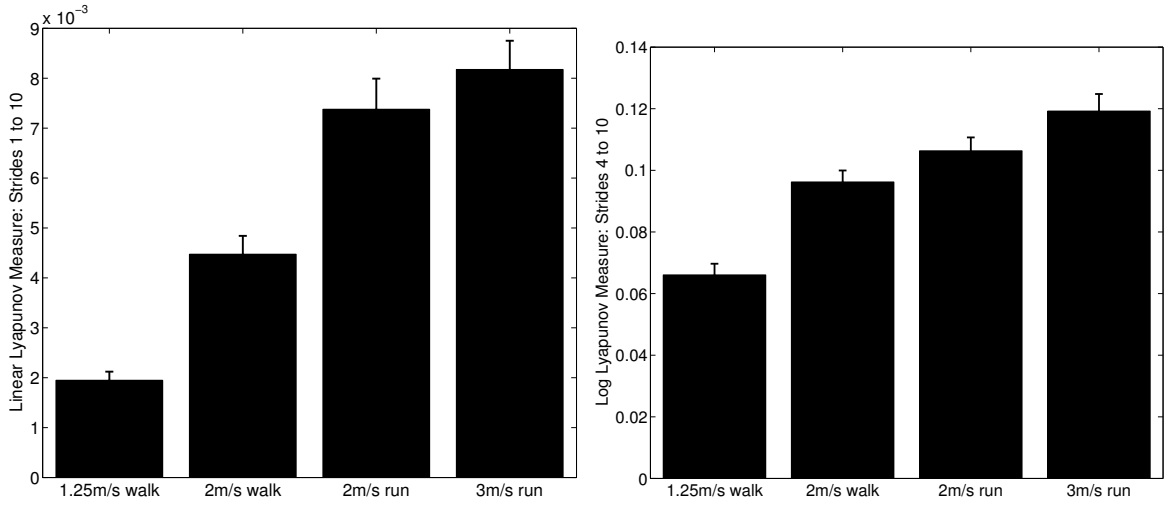

Figure 3.4: Lyapunov measures for vertical position of the sacrum.

The Lyapunov measure changed distinctly with condition. There was an increase from walking at 1.25m/s walking to 2m/s walking, and another increase from 2m/s walking to 2m/s running. Additionally, the value for 3 m/s running was greater than that for 2 m/s running. Figure shows means ± SEM.

interaction effect ($p = 0.00082$; Figure 3.6). The Lyapunov measure values for walking at

1.25 m/s were significantly less than those found in walking at 2 m/s ($p < 0.01$). The values found for running at 2 m/s were smaller than those found for walking at 2 m/s ($p < 0.01$). However, values for running at 3 m/s were significantly greater than those found for 2 m/s ($p < 0.01$). Though these values come from calculating the slope for the first 10 strides for the non-logged data, similar results were found whether we used strides 1-5, strides, 4-10, or strides 1-10 to calculate these values. Furthermore, we found similar results regardless of whether or not we log transformed the data to calculate the Lyapunov exponent.

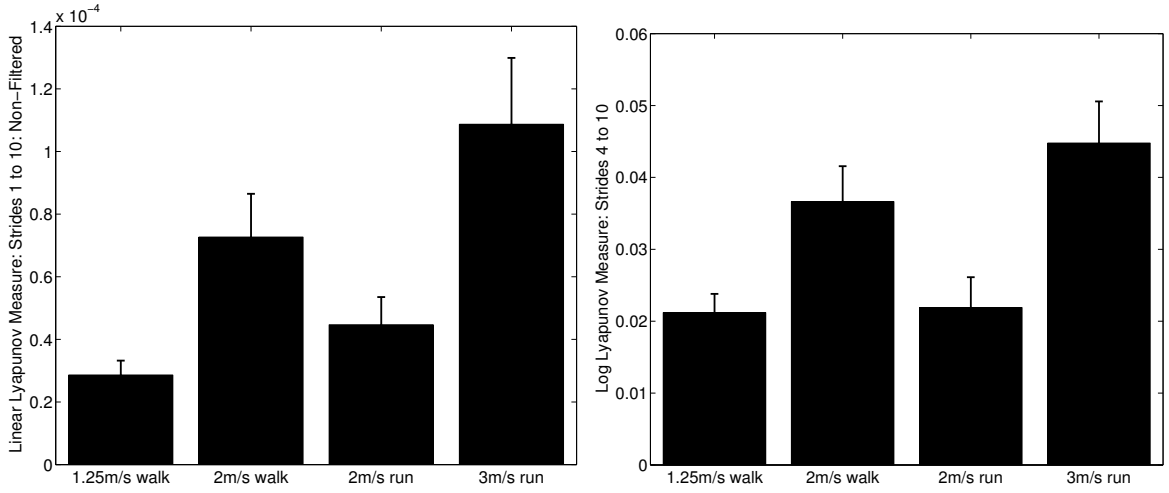

Figure 3.5: Lyapunov measures for mediolateral position of the sacrum.

The Lyapunov measures for the mediolateral position of the sacrum showed a general pattern of increase of Lyapunov measure with increase with speed, regardless of gait. The Lyapunov measure values for walking at 1.25 m/s were greater than those for walking at 2 m/s. The Lyapunov measure values for running at 2 m/s were smaller than those for walking at 2 m/s. Furthermore, the Lyapunov measure values for running at 3 m/s were greater than those for running at 2 m/s. Figure shows means ± SEM.

Vertical Position of the Toe Using the vertical position of the toe, we found a significant effect of speed ($p = 0.038$), as well as a significant interaction effect ($p = 6.6 \times 10^{-6}$; Figure 3.6). The Lyapunov measure values for walking at 1.25 m/s were significantly lower than those found in walking at 2 m/s ($p < 0.01$). The values found in running at 2 m/s were smaller than those found for walking at 2 m/s ($p < 0.01$). However, values for running

at 3 m/s were significantly greater than those found for 2 m/s ($p < 0.01$). Those these values come from calculating the slope of the non-logged data for the first 10 strides, similar results were found whether we used strides 1-5, strides, 4-10, or strides 1-10 to calculate these values. Furthermore, we found similar results regardless of whether or not we took log of the data to calculate the Lyapunov exponent.

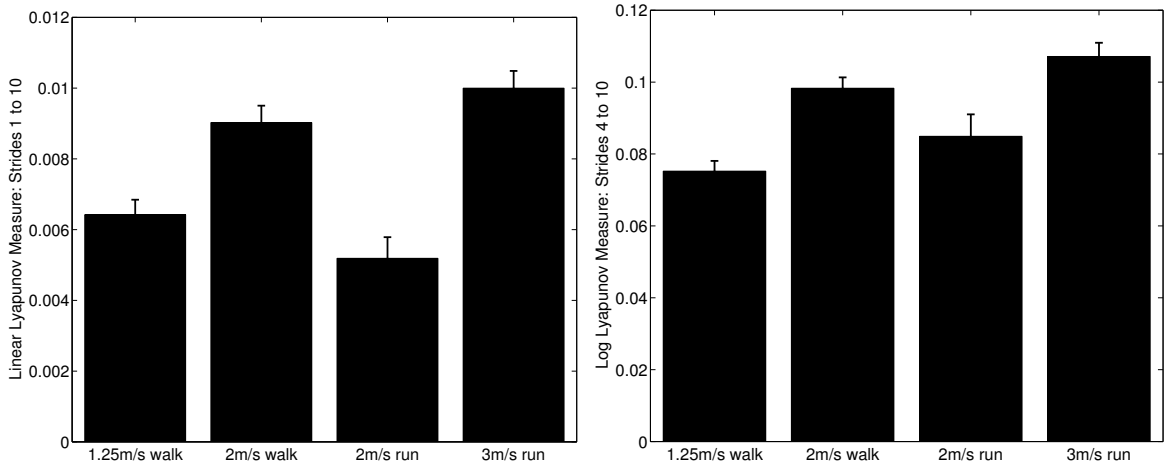

Figure 3.6: Lyapunov measures for the vertical position of the toe.

The Lyapunov measures for the vertical position of the left toe showed a general pattern of increase of Lyapunov measure with increase with speed, regardless of gait. The Lyapunov measure values for walking at 2 m/s were greater than those for walking at 1.25 m/s. The Lyapunov measure values for running at 2 m/s were smaller than those for walking at 2 m/s. Furthermore, the Lyapunov measure values for running at 3 m/s were greater than those for running at 2 m/s. Figure shows means ± SEM.

3.3.3 Time and Strides to Saturation

For the vertical position of the sacrum, we found no significant differences between conditions for time or strides to saturation (Figure 3.7). Though the trends change for time as opposed to strides, there are no significant differences for either measure.

However, we did find differences in time to saturation for the mediolateral position of the sacrum (Figure 3.8). We found a significant speed effect ($p = 0.0015$). However, this effect was no longer present for strides to saturation.

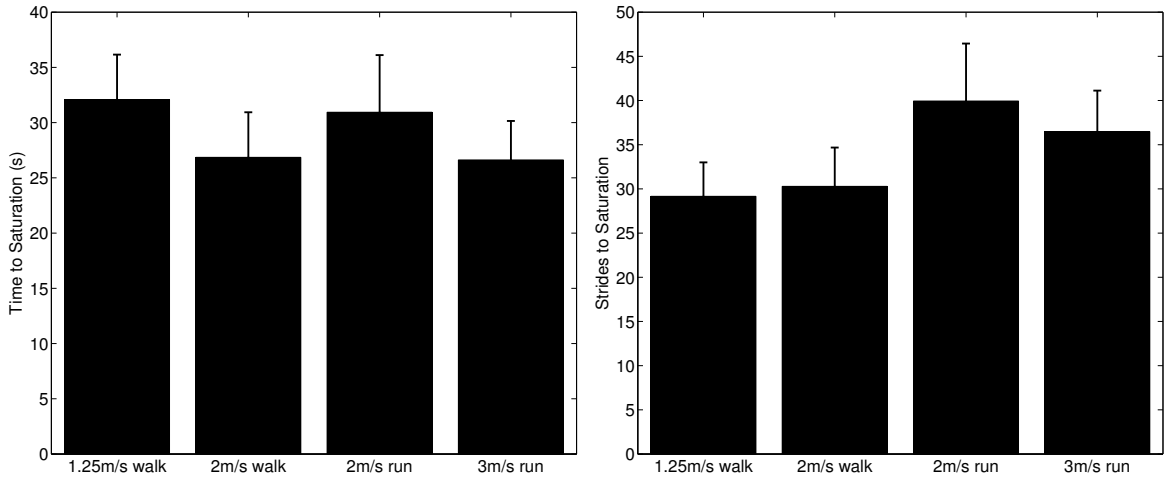

Figure 3.7: Time and number of strides to saturation for the vertical position of the sacrum

We find no significant differences between the time or strides to saturation for the vertical position of the sacrum. Figure shows means ± SEM.

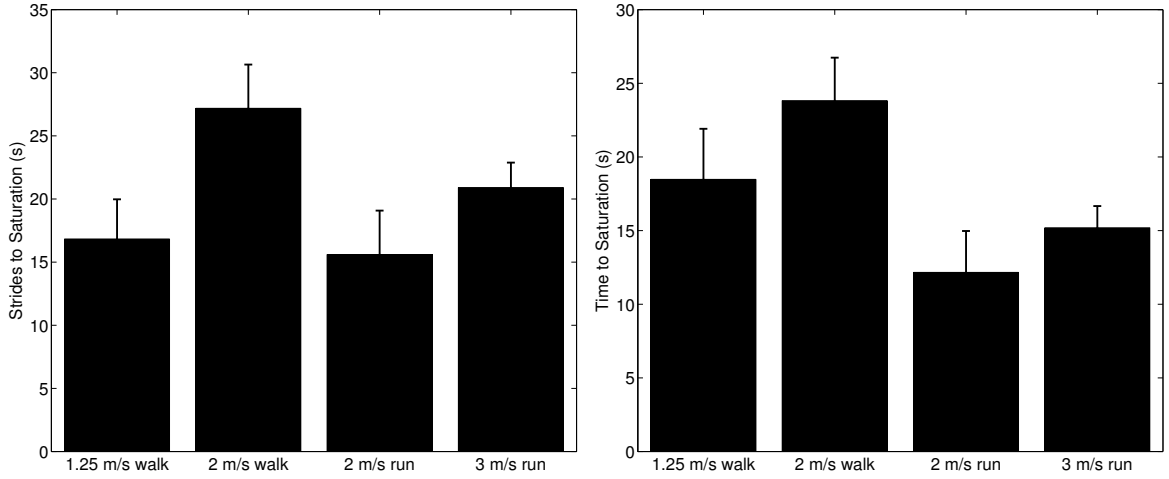

Figure 3.8: Time and number of strides to saturation for the mediolateral position of the sacrum

We found the same general pattern in time and strides to saturation for the mediolateral sacrum as we found for the Lyapunov measures for these values, but statistically, there was only an effect of speed. Figure shows means ± SEM.

For the vertical position of the toe, we found no significant differences between conditions for time to saturation either in terms of time or in terms of strides (Figure 3.9).

Filtering had a significant effect on the data, leading to higher Lyapunov measure

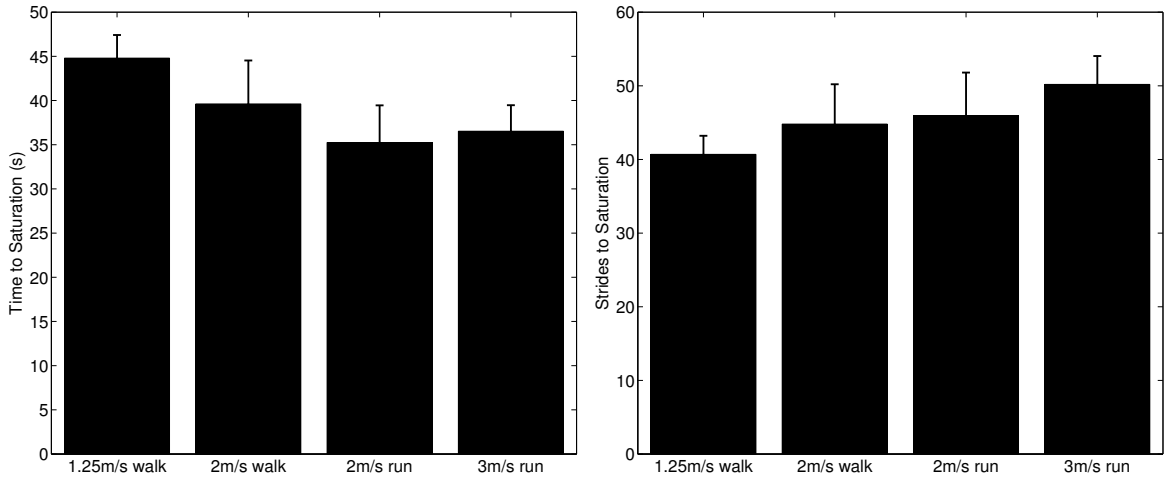

Figure 3.9: Time and number of strides to saturation for the vertical position of the toe.

We found no significant differences between the time or strides to saturation for the vertical position of the toe. Figure shows means ± SEM.

values than without filtering (Figure 3.10). Similarly, as one would expect, an increase in embedding dimension led to an increase in the value of the Lyapunov exponent. These results for a typical subject are shown in Figures 3.11 and 3.13.

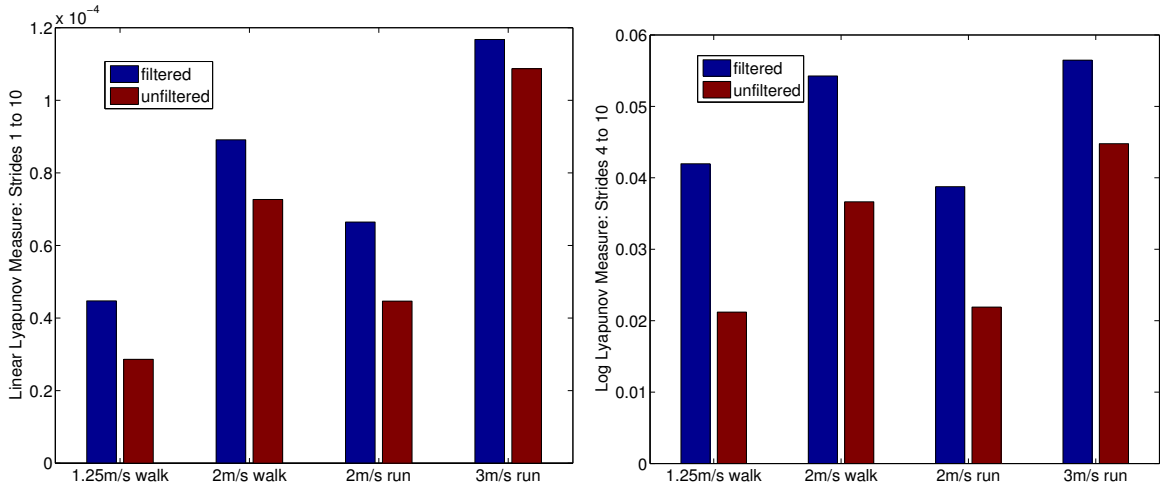

Figure 3.10: Effect of filtering on Lyapunov measure.

As is shown for this data for all subjects, we found a general increase in the Lyapunov exponent found for filtered as opposed to unfiltered data. However, the net increase in value due to filtering was not consistent across conditions.

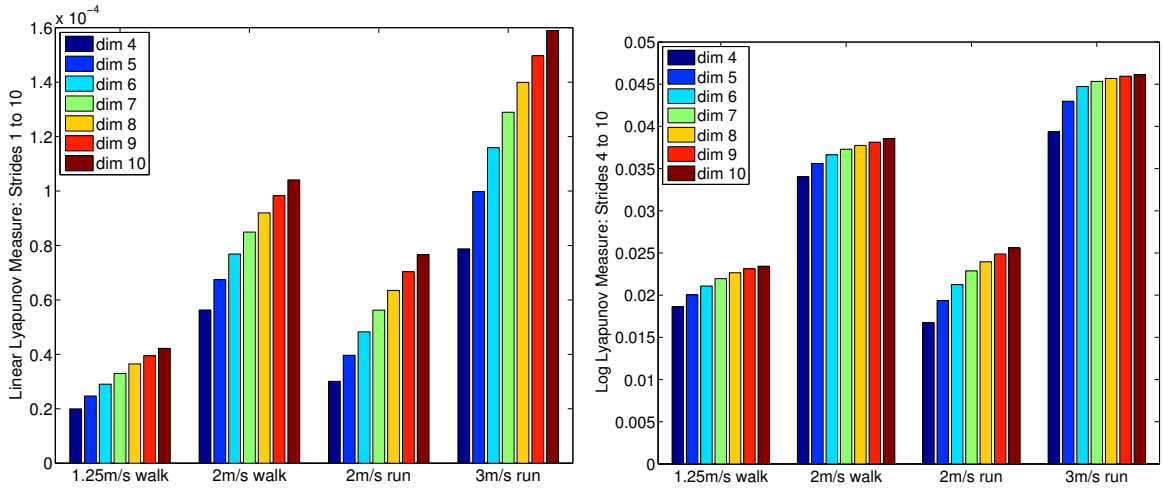

Figure 3.11: Effect of embedding dimension on unfiltered data.

For filtered data, an increase in embedding dimension leads to an increase in the Lyapunov exponent, with larger increases found for smaller embedding dimensions than larger ones. The data shown is the average value for all subjects for all conditions.

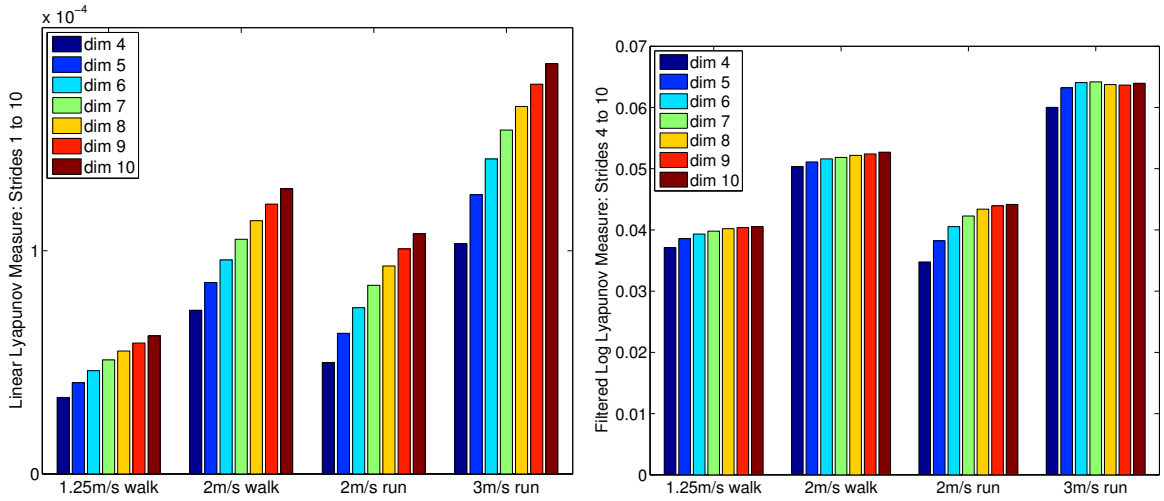

Figure 3.12: Effect of embedding dimension on filtered data.

For unfiltered data, an increase in embedding dimension leads to an increase in the Lyapunov exponent, with larger increases found for smaller embedding dimensions than larger ones. The data shown is the average value for all subjects for all conditions.

3.3.4 Changes Across Transition

Our initial results indicate an increase in stability for the first 100 strides after the transition point. For heelstrike, we find a gradual decrease in eigenvalue from the transition

forward. We found a decrease whether the transition was from walking to running ($p = 0.0025$) or from running to walking ($p = 0.0029$). Further research will have to be done to determine if this pattern holds across stance and across a stride.

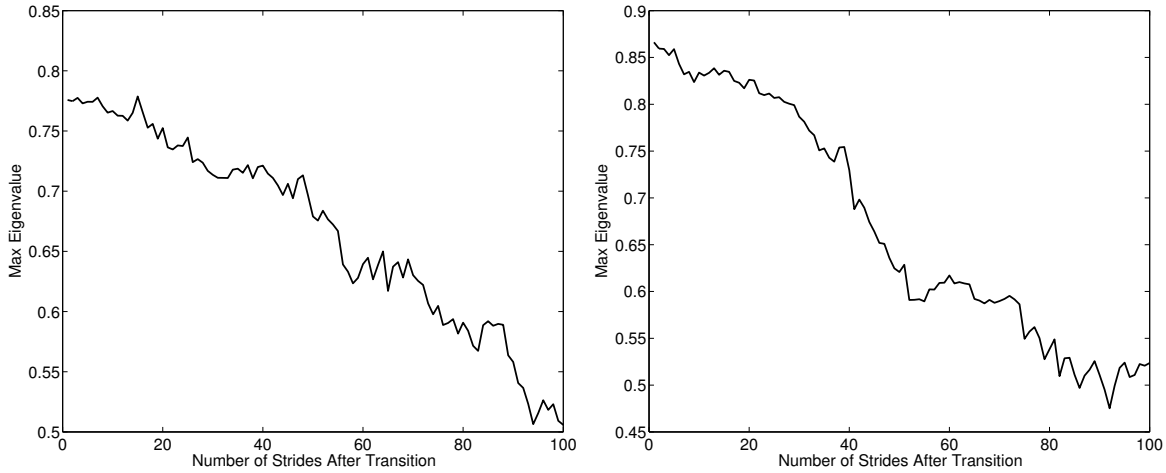

Figure 3.13: Changing stability immediately after gait transition.

For both the walk-to-run transition and the run-to-walk transition, there is a decrease in the maximum eigenvalue of joint angles and angular velocities over the course of the first 100 strides after a change in gait.

3.4 Discussion

We found different results for the stability of the system, depending on the measure used, making it difficult to conclude anything about the stability of the overall system across gait. We found similar patterns in the eigenvalues of the joint angles and angular velocities and the Lyapunov measures of the vertical position of the sacrum. On its own, our data suggest that the vertical measure of the sacrum is enough to capture the dynamics of the system in the sagittal plane. The pattern found in both of these measures indicates a significant decrease in stability in walking with an increase in speed. Further, running at the approximate walk-run transition speed is less stable than walking at the same speed. These data suggest that running is less stable than walking, but that stability in running does not

depend on speed. Stability in walking decreases at faster speeds, but remains more stable than running even at transition speed. These data matches the sagittal plane data from the passive dynamic walkers (McGeer, 1990a,b). The change across gait is also similar to the results found for the head by (Jordan et al., 2008). It should also be noted that we found similar embedding dimensions as Jordan et al. (2008). They found an embedding dimension of d=5 to be appropriate for all trials, whereas for our data, it varied from 5 to 6, depending on the trial.

We found similar values when we examined the Lyapunov exponents associated with the mediolateral sacrum position and the vertical position of the toe. However, the patterns for both these measures differed from the results found from the eigenvalue analysis and the Lyapunov analysis of the vertical position of the sacrum. In this case, we found a significant increase in the Lyapunov measure as speed increased for each gait. The Lyapunov measure for walking at 2 m/s is much greater than for walking at 1.25 m/s. Similarly, the Lyapunov measure for 3 m/s running is much greater than that for 2 m/s running. However, we found that running at 2 m/s appears to be much more stable than walking at the same speed. Because Jordan et al. (2008) only examined the vertical direction of the head marker, we cannot compare our data with their results.

The only consistent result for all of the measures and directions examined is that walking stability decreases from a typical walking speed to approximate transition speed. This is consistent with previous findings using these measures that show a general increase in stability past preferred walking speed (Dingwell and Cusumano, 2000; Dingwell et al., 2000, 2001).

The evidence for changes in stability is further complicated by the time and strides to saturation data. We found no differences in the time or strides to saturation for either of the vertical position measures, the sacrum or the toe, but we did find a difference in the mediolateral sacrum time to saturation. This may indicate that the attractor for the mediolateral position of the sacrum is larger for the faster walking and running speeds, and

this is the reason we found the larger magnitude Lyapunov measures.

It has been argued that using time-delayed copies of one specific measure should be sufficient to elucidate the dynamics of the entire system, but our data suggest otherwise. We do find that some measures give distinctly similar results. The vertical position of the sacrum shows a similar pattern to the joint flexion extension velocity state-space analysis. The mediolateral position of the sacrum also shows a similar pattern to the vertical position of the toe. However, according to the theory underlying the use of Lyapunov measurements, if the markers are really elucidating the underlying dynamics of the system, they should all give similar results. This suggests that different markers may be imparting different information about the state of the system. It is possible that the vertical position of the toe and the mediolateral position of the sacrum are capturing some critical dynamics of the system across transition, as these measures show increasing stability with going from walking to running at 2 m/s.

Taken as a whole, our results suggest that a great deal of caution should be used in both the use and interpretation of any nonlinear dynamics measurements. We found different results depending on not just which marker is analyzed, but the direction in which the marker position measurement was taken. Past studies have not been consistent in terms of what measure has been used in order to produce the Lyapunov measure (Dingwell and Cusumano, 2000; Dingwell et al., 2000, 2001; Jordan et al., 2008). Also, many common assumptions that are made in nonlinear dynamic measures can affect the values of the data. Though it has been suggested that the data should not be filtered, some researchers filter the data (Jordan et al., 2008) , whereas others do not (Dingwell and Cusumano, 2000). Filtering can potentially change the embedding dimension identified. Furthermore, the embedding dimension used can significantly affect the value of the Lyapunov measure, regardless of whether the log is taken or not. Some studies individually identify the appropriate embedding dimension for each trial, as we did, whereas others just find one embedding dimension and assume they will be consistent for all measures. Our sensitivity analysis indicate that these two strategies

can give different results. Lastly, detrending the data reduces the size of the attractor, and with motion capture data, often increases the noise to signal ratio. This sensitivity makes it still more difficult to identify an accurate embedding dimension. Similarly, studies do not always use the same process for identifying the Lyapunov exponent. Often, they use various methods to identify a "linear" region in the data. However, we did not find a distinct linear region in our data for any of the markers analyzed, which suggests that this type of analysis is problematic. Overall, because there is no consistency in the methodology, and because the value of the "exponent" varies widely depending on the methodology used, it is impossible to compare the values of Lyapunov exponents across different studies.

If taken at face value, our results indicate that stability decreases with increasing walking speed. Additionally, stability of some measures decreases when subjects walk as opposed to run at the same speed, but stability of other measures decreases when subjects walk as opposed to run at the same speed. Further, within running, whether or not stability increases or remains constant with speed also depends on the measure being examined. However, our results indicate some significant troubles with using the methodology outlined for calculating Lyapunov exponents, indicating that it may not be measuring what it has been assumed to measure. It is possible that the quantity being measured still has some physiological meaning, but other, more sophisticated methods should be used to determine if a Lyapunov analysis is appropriate for gait data, and further, what this value is truly quantifying. Further analysis will use software such as TISEAN (Hegger et al., 1999) to compare this analysis with one that allows for more parameter adjustment in order to determine whether the method used here is too simple to capture the true underyling system, or whether what is being measured is merely added noise. If this analysis proves appropriate, we will compare these results with recently introduced measures of stability (Hof et al., 2005).

Chapter 4

Using Projection and DMOC to find optimal control in systems with and without impacts

Collaborator: Todd D. Murphey, Northwestern University

4.1 Introduction and Background

Discrete Mechanics and Optimal Control, abbreviated DMOC, is a method that uses variational integrators to find the optimal control for a given discretized physical system using ideas from Lagrangian and Hamiltonian dynamics (Marsden and West, 2003; Ober-Bloebaum et al., 2008). This enables DMOC to be effective for impacts and collisions as well as smooth motion in both linear and non-linear systems. Rather than discretizing a continuous equation, DMOC uses the forced discrete Euler-Lagrange (DEL) equations over multiple time steps as constraints and, within these constraints, minimizes a cost functional of weighted discrete control values. It avoids many of the problems that accompany discretizing continuous equations of motion, such as inaccurate loss or gain of energy over time.

The ability of DMOC to handle both non-linearity and impacts while retaining the essential characteristics of the underlying physical system has made it useful in a variety of arenas. These include undercontrolled systems, such as a group of hovercraft that have

control in only two of their three degrees of freedom (Junge et al., 2005) and constrained multibody dynamcs (Leyendecker et al., 2007). DMOC has also been used as a way to find optimal control in hybrid systems, such as bipedal walking in two dimensions (Pekarek et al., 2007).

DMOC's effectiveness, however, depends on the method used to solve the ultimate constrained optimization problem. Traditionally, DMOC has been solved using Sequential Quadratic Programming (SQP) (Marsden and West, 2003; Pekarek and Marsden, 2008). SQP is a method that combines the cost function and the linearized constraints into an approximation of the Lagrangian function. At each step, it solves a quadratic programming sub-problem based on this approximation, then updates the approximation of the Hessian for the next iteration. Though this method is effective, because it is essentially a 'black box' algorithm for any arbitrary non-linear constrained optimization problem, it typically ignores some of the analytic properties of DMOC that could be used to improve convergence.

We present a method that uses the structure of the equations involved in DMOC to build a projection of arbitrary configuration and control variables onto feasible configuration and control variables. This projection can then be used to reduce the dimensionality of the problem via the constraints, allowing the optimal solution to be found via pure Newton's method without losing any information. We then present preliminary findings on how the projection method compares to both SQP with numerical derivative calculations and gradient descent with Armijo line search. The three methods are compared in terms of accuracy of results and convergence properties for a variety of example problems.

We organize the paper in the following way: Section II provides an explanation of how the continuous system can be modeled using discrete mechanics, while retaining the physical characteristics of the original problem. Section III gives a short background on the structure and use of DMOC. Section IV provides a description of our method, including the derivation of the projection and its relevant derivatives. Section V compares examples of our method to both gradient descent and numerical SQP for two non-linear systems. Section VI gives

some very preliminary work on a simple system involving impacts. Lastly, Section VII gives conclusions and future work.

4.2 Discrete Mechanics

We begin with a continuous mechanical system with trajectory $q(t) \in Q$ (configuration space) starting at some initial point $(q(0), \dot{q}(0))$and ending at $(q(T), \dot{q}(T))$. The system is also subject to some external forcing, $u(t)$, with the optimal control determined by minimizing the cost functional

$$J(q, u) = \int_0^T C(q(t), \dot{q}(t), u(t))dt. \tag{4.1}$$

To preserve the mechanics of the original system, the system is constrained by the Lagrange-D'Alembert principle

$$\delta \int_0^T L(q(t), \dot{q}(t))dt + \int_0^T u(t)\delta q(t)dt = 0 \tag{4.2}$$

for all variations $\delta q(t)$ s.t. $\delta q(0) = \delta q(T) = 0$, where L is the Lagrangian mapping the tangent bundle of Q, TQ, to $\mathbb{R}$, $L : TQ \to \mathbb{R}$.

The above two equations describe a continuous constrained optimization problem; to perform DMOC, we need an approximation of this continuous problem in discrete space. A complete discussion and derivation of this discretization can be found in (Ober-Bloebaum et al., 2008), whereas much of the following shorter derivation is taken from (Marsden and West, 2003) and (Junge et al., 2005).

To define the discrete problem, we first convert the state space from TQ as $Q \times Q$ by replacing a pair $(q, \dot{q})$ with the triplet (q_0, q_1, h) where h is the timestep (for more details see (Marsden and West, 2003)). We then discretize the continuous path $q : [0, T] \to Q$ via $q_d : \{0, h, 2h, \ldots, Nh\} \to Q$, where $N \in \mathbb{N}$ and $T = Nh$, and we approximate $q(kh)$ with $q_k = q_d(kh)$. Similarly, we transform the continuous control $u : [0, T] \to T^*Q$ (the cotangent bundle) to the discretized values $u_d : \{0, h, 2h, \ldots, Nh\} \to T^*Q$, where again, $u_k = u_d(kh)$.

We next need the discretized version of the continuous cost functional and Lagrange-D'Alembert principle. We first discretize the Lagrangian L, giving

$$L_d(q_k, q_{k+1}) \approx \int_{kh}^{(k+1)h} L(q(t), \dot{q}(t))dt \tag{4.3}$$

$$\approx hL(\frac{q_{k+1}+q_k}{2}, \frac{q_{k+1}-q_k}{h}), \tag{4.4}$$

with the midpoint rule used in the discretization. Similarly, we approximate a discretized version of the control variable:

$$u_k^- \cdot \delta q_k + u_k^+ \cdot \delta q_{k+1} \approx \int_{kh}^{(k+1)h} u(t) \cdot \delta q(t)dt. \tag{4.5}$$

$$\tag{4.6}$$

where u_k^+ and u_k^- represent the right-hand and left-hand control for the time step k. Using the discrete Lagrangian, we can define the discrete Lagrange-D'Alembert principle, which restricts to paths $\{q_k\}_{k=0}^N$ such that all variations $\{\delta q_k\}_{k=0}^N$ with $\delta q_0 = \delta q_N = 0$ require that

$$\delta \sum_{k=0}^{n-1} L_d(q_k, q_{k+1}) + \sum_{k=0}^{n-1} u_k^- \cdot \delta q_k + u_k^+ \cdot \delta q_{k+1} = 0. \tag{4.7}$$

Rewriting, we get the forced discrete Euler-Lagrange (DEL) equations:

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) + u_{k-1}^+ + u_k^- = 0. \tag{4.8}$$

where D_i represents the derivative with respect to the i^{th} argument (i.e. $D_2 L_d(q_{k-1}, q_k)$ is the derivative of L_d with respect to q_k) for $k = 1$ to N. Lastly, for each time step k, the cost functional can be rewritten as

$$C_d(q_k, q_{k+1}, u_k, u_{k+1}) \approx \int_{kh}^{(k+1)h} C(q, \dot{q}, u)dt \tag{4.9}$$

which, leads to the total cost:

$$J_d(q_d, u_d) = \sum_{k=0}^{N-1} C_d(q_k, q_{k+1}, u_k, u_{k+1}). \tag{4.10}$$

4.3 DMOC

The equations for all interior time steps can be described by (4.8), but we must also ensure these interior points are continuous with the fixed initial and final time steps to meet these boundary conditions. To do so, the discrete Legendre transforms $\mathbb{F}^+$ and $\mathbb{F}^-$ are used to relate the representation in the continuous (TQ) and discrete ($Q \times Q$) domains. These transforms are defined to be

$$\mathbb{F}^+ L_d \quad : (q_{k-1}, q_k) \to (q_k, p_k) \tag{4.11}$$

$$p_k = D_2 L_d(q_{k-1}, q_k) + u^+_{k-1} \tag{4.12}$$

$$\mathbb{F}^- L_d \quad : (q_{k-1}, q_k) \to (q_{k-1}, p_{k-1}) \tag{4.13}$$

$$p_k = -D_1 L_d(q_{k-1}, q_k) - u^-_{k-1}. \tag{4.14}$$

Further descriptions of these transforms can be found in (Marsden and West, 2003). Additionally, the standard Legendre transform can be used to map TQ to T^*Q.

$$\mathbb{F}L : (q, \dot{q}) \to (q, p) = (q, D_2 L(q, \dot{q})) \tag{4.15}$$

This gives the following additional constraint equations for the initial ($t = 0$) and final ($t = T$) time steps.

$$D_2 L(q(0), \dot{q}(0)) + D_1 L_d(q_0, q_1) + u^-_0 = 0 \tag{4.16}$$

$$-D_2 L(q(T), \dot{q}(T)) + D_1 L_d(q_{n-1}, q_n) + u^+_{n-1} = 0. \tag{4.17}$$

Thus our final system consists of minimizing the discrete cost functional (4.10) subject to the constraints (4.8), (4.16) and (4.17).

In traditional DMOC, both u^-_k and u^+_k are calculated via the midpoint rule as $\frac{h}{4}(u_k + u_{k+1})$. However, to make the projection simpler in the next section, we use a slightly different approximation, using the left hand endpoint rather than the midpoint rule. Instead of averaging the two controls, we simply denote both u^-_k and u^+_k as u_k. Note that this does

not significantly change the optimal control, and that midpoint control values can easily be recovered using the previous definition of u_k^-.

We thus have the overall DEL equations

$$D_2L(q(0),\dot{q}(0)) + D_1L_d(q_0,q_1) + u_0 = 0 \tag{4.18}$$

$$D_2L_d(q_{k-1},q_k) + D_1L_d(q_k,q_{k+1}) + u_{k-1} + u_k = 0 \tag{4.19}$$

$$-D_2L(q(T),\dot{q}(T)) + D_1L_d(q_{n-1},q_n) + u_{n-1} = 0. \tag{4.20}$$

4.4 Projection-Based Optimization

DMOC is traditionally formulated as a constrained optimization problem. For this reason, the problem is generally framed as trying to find the trajectory and control that minimize the cost functional within a subset of the domain that obeys the constraints

$$\min_{(q,u)\in U\subseteq Q\times T^*Q} J(q,u).$$

In this formulation, the cost is computed while assuring the trajectory and control variables meet the constraints. Thus, though the search stays near the constraint surface, in theory, the search is being done over the entire set of trajectory and control values, not just those subject to the constraints.

It is equivalent to define the unconstrained minimization problem of finding the minimal control over configuration and control variables constrained via a projection $\mathscr{P}$,

$$\min_{x\in B_\epsilon(q,u))} J(\mathscr{P}(q,u)),$$

where x is a combination of possible configuration and control variables and $B_\epsilon(q,u)$ is a neighborhood of (q,u). In practice, this search can be done by projecting a given input of arbitrary configuration and control variables onto the subspace of admissible configurations and controls and composing that projection with the cost. Thus rather than trying to simultaneously minimize J while fulfilling the constraints, only the subspace of admissible

configurations and controls is ever examined. We therefore use the projection to reduce the dimensionality of the problem via the constraints before ever attempting to find the minimum.

However, the problem now hinges on the choice of the projection $\mathscr{P}$. $\mathscr{P}$ must be defined so that it takes any arbitrary combination of configuration and control variables and maps it onto one that fulfills all of the constraint equations. Furthermore, in order to optimize over all possible projections, we will also need both first and second derivatives of $\mathscr{P}$, $D\mathscr{P}$ and $D^2\mathscr{P}$, thus requiring $\mathscr{P}$ to be at least C^2 in all its arguments.

4.4.1 Projection Definition and Differentiability: No Impact

The DEL equations (4.18), (4.19) and (4.20) can be used to define the projection $\mathscr{P}$ and its derivatives for an arbitrary DMOC problem. Recall that the initial boundary condition is

$$D_2L(q(0),\dot{q}(0)) + D_1L_d(q_0,q_1) + u_0 = 0. \tag{4.21}$$

Defining

$$g_0(q(0),\dot{q}(0),q_0,q_1) = D_2L(q(0),\dot{q}(0)) + D_1L_d(q_0,q_1), \tag{4.22}$$

we can rewrite the DEL equation as

$$g_0(q(0),\dot{q}(0),q_0,q_1) + u_0 = 0. \tag{4.23}$$

We then solve for u_0 as a function of the boundary conditions $q(0) = q_0$ and $\dot{q}(0)$ and the configuration variable q_1, giving

$$u_0 = -g_0(q(0),\dot{q}(0),q_0,q_1). \tag{4.24}$$

Any DEL equation for an interior point $k = 1,\ldots,n-1$

$$D_2L_d(q_{k-1},q_k) + D_1L_d(q_k,q_{k+1}) + u_{k-1} + u_k = 0, \tag{4.25}$$

can also be rewritten as

$$g(q_{k-1}, q_k, q_{k+1}) + u_{k-1} + u_k = 0, \tag{4.26}$$

where

$$g(q_{k-1}, q_k, q_{k+1}) = D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}). \tag{4.27}$$

Solving for u_k, we have the feedback law

$$u_k = -u_{k-1} - g(q_{k-1}, q_k, q_{k+1}). \tag{4.28}$$

After solving the first equation for u_0, the remaining interior equations can be solved for the control variables in terms of the configuration variables for $k = 1, \ldots n-1$, giving

$$u_k = (-1)^{k+1} g_0(q(0), \dot{q}(0), q_0, q_1) + \sum_{j=1}^{k} (-1)^j g(q_{j-1}, q_j, q_{j+1}). \tag{4.29}$$

This defines a projection $\mathscr{P}(q, u) = (q, u(q))$ on the space of all configuration and control variables, where each u_i is defined via (4.24) and (4.29). u does not appear in the projection because it can be fully defined from the configuration variables, so $u(q)$ does not depend on u.

Whether the projection becomes more complex depends on how the ending condition is enforced, which can be done in two different ways. One way is to impose a terminal cost by including a term in the cost function enforcing a steep cost penalty for deviating from the final condition. Because this method of prescribing the ending condition requires no further constraints, and because we have thus far used the n constraints to solve for n of the $2n$ total variables in the system, this completes the projection.

Alternatively, imposing the ending condition using the remaining DEL equation gives n constraints and $2n-1$ degrees of freedom, requiring one of the configuration time steps to be solved via the others. This can be done by again using the DEL equations (4.19) and (4.20). We first rewrite the DEL equation (4.20) for the ending condition as

$$g_n(q(T), \dot{q}(T), q_{n-1}, q_n) + u_n = 0 \tag{4.30}$$

where

$$g_n(q(T), \dot{q}(T), q_{n-1}, q_n) = -D_2L(q(T), \dot{q}(T)) + \tag{4.31}$$

$$D_1L_d(q_{n-1}, q_n). \tag{4.32}$$

If we denote (4.18) DEL_0 and (4.19) DEL_k for $k = 1 \ldots n-1$ we can combine them to give

$$0 = \sum_{i=0}^{n}(-1)^i DEL_i \tag{4.33}$$

$$0 = g_0(q(0), \dot{q}(0), q_1) + (-1)^{n+1}g_n(q_n, \dot{q}_n, q_{n-1}) + \tag{4.34}$$

$$\sum_{i=1}^{n-1}(-1)^i g(q_{i+1}, q_i, q_{i-1}) \tag{4.35}$$

which depends only on the configuration variables. This equation can be used to, usually implicitly, solve for either the first or last time step in terms of the configuration variables. Note that because q_n is a boundary condition, our last configuration variable in this case is actually q_{n-1}.

Using these equations, we can construct an appropriate projection $\mathscr{P}(q, u)$. If we include the ending condition in the cost equation, we have a projection of our configuration variables onto themselves, and the control as a function of our configuration variables. Thus, our projection becomes

$$\mathscr{P}(q, u) = (q, u(q)). \tag{4.36}$$

If the DEL equation is used to impose the ending condition, we get the projection

$$\mathscr{P}(q, u) = (q_1(q_2, \ldots, q_{n-1}), q_2, \ldots, q_{n-1}, \tag{4.37}$$

$$u_1(q_2, \ldots, q_{n-1}), \ldots, u_n(q_2, \ldots, q_{n-1})). \tag{4.38}$$

Lemma 1. *$\mathscr{P}$ is a projection, and $L_d, L \in C^n \rightarrow \mathscr{P} \in C^{n-1}$*

Proof. For the case of terminal cost,

$$\mathscr{P}(\mathscr{P}(q, u)) = \mathscr{P}(q, u(q)) = (q, u(q)) = \mathscr{P}(q, u). \tag{4.39}$$

If the ending condition is met by including another constraint equation

$$\begin{aligned}
\mathscr{P}(\mathscr{P}(q,u)) &= \mathscr{P}(q_1(q_2,\ldots,q_{n-1}),q_2,\ldots,q_{n-1}, && (4.40)\\
&\quad u_1(q_2,\ldots,q_{n-1}),\ldots,u_n(q_2,\ldots,q_{n-1})) && (4.41)\\
&= (q_1(q_2,\ldots,q_{n-1}),q_2,\ldots,q_{n-1}, && (4.42)\\
&\quad u_1(q_2,\ldots,q_{n-1}),\ldots,u_n(q_2,\ldots,q_{n-1})) && (4.43)\\
&= \mathscr{P}(q,u). && (4.44)
\end{aligned}$$

Furthermore, because $\mathscr{P}$ depends solely upon the DEL equations, if $L_d, L \in C^n$, since $\mathscr{P}$ consists of elements which are linear combinations of DL_d and DL, then we must have $\mathscr{P} \in C^{n-1}$. □

To use traditional unconstrained optimization algorithms, such as Newton's method, requires the derivatives of the constrained cost function. These derivatives can be found using

$$DJ_c = DJ(\mathscr{P}(q)) \cdot D\mathscr{P}(q), \tag{4.45}$$

where J denotes the unconstrained cost function and J_c the cost function constrained via the projection. Similarly, the Hessian can be calculated via

$$\begin{aligned}
D^2J_c &= D^2J(\mathscr{P}(q)) \cdot D\mathscr{P}(q) + && (4.46)\\
&\quad DJ(\mathscr{P}(q)) \cdot D^2\mathscr{P}(q) && (4.47)
\end{aligned}$$

where the elements of $D\mathscr{P}$ and $D^2\mathscr{P}$ can be found by taking the first and second derivatives the projection elements as defined via the DEL equations.

We then use Newton's method to find the minimum cost of the projected system.

4.5 Examples

4.5.1 Incremental Cost

For the example of the double pendulum, the projection method was compared to two other methods, numerical SQP and gradient descent using Armijo line search (Kelley, 1995). For all tests, SQP was implemented using MATLAB's fmincon function using numerical derivatives with constraint tolerance set to 10^{-14}. SQP takes a constrained optimization problem that is trying to minimize the value of $f(x)$ subject to the constraints $g(x) = 0$, and rewrites the problem as trying to minimize $L(x) = f(x) - \lambda c(x)$, where L is the Lagrangian, and λ are the Lagrange multipliers (Conn et al., 2000; Nocedal and Wright, 1999). This is implemented by, at each step, solving the quadratic subproblem

$$1/2p \cdot \Delta_{xx} L(x, \lambda)p + \Delta_x L(x, \lambda) \cdot p + L(x, \lambda) \tag{4.48}$$

subject to

$$\Delta_x c(x)^T p + c(x) = 0. \tag{4.49}$$

This can be formulated as Newton's method on the system

$$\Delta_x f(x) - \lambda \Delta_x c(x) = 0 \tag{4.50}$$

$$-c(x) = 0 \tag{4.51}$$

essentially solving for the ideal x and λ values at once. This can be implemented by solving

$$\begin{bmatrix} \Delta_{x_i x_i} L(x_i, \lambda_i) & -\Delta c(x^k) \\ -\Delta c(x^k)^T & 0 \end{bmatrix} \begin{bmatrix} p_i \\ \phi_i \end{bmatrix} = \begin{bmatrix} \Delta L(x_i, \lambda_i) \\ -c(x_i) \end{bmatrix} \tag{4.52}$$

for p_i and ϕ_i, and updating x_i and λ_i via $(x_{k+1}, \lambda_{k+1}) = (x_i, \lambda_i) + (p_i, \phi_i)$. In practice, the Hessian can be user-defined or estimated for each step. There are, however, some complexities introduced in how the Hessian ($\Delta_{x_i x_i} L(x_i, \lambda_i)$) is estimated and/or updated (Nocedal and Wright, 1999).

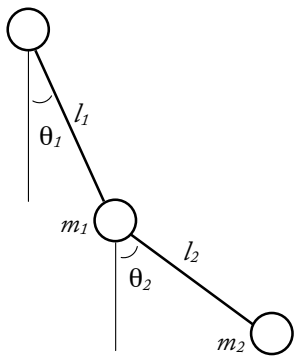

Figure 4.1: A depiction of the double pendulum system, including definitions of the variables θ_1 and θ_2 and the parameters m_1, m_2, l_1 and l_2.

For this initial case, we used a cost function of

$$J(q_i, u_i) = \sum_{i=0}^{n-1} u_i^T u_i \tag{4.53}$$

where u_i was defined in (4.24) and (4.29).

For our simulations, θ_1 represents the angle with respect to the vertical at the base joint and θ_2 represents the angle with respect to the vertical at the outer joint (Figure 4.1). The Lagrangian of the double pendulum is

$$\begin{aligned} L &= \frac{1}{2}(m_1 + m_2) l_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 l_2^2 \dot{\theta}_2^2 && (4.54) \\ &\quad + m_2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \cos(\theta_1 - \theta_2) + && (4.55) \\ &\quad (m_1 + m_2) g l_1 \cos(\theta_1) + m_2 g l_2 \cos(\theta_2) && (4.56) \end{aligned}$$

where m_1 and m_2 are the masses for the respective joints, and l_1 and l_2 are the lengths of the segments from the base joint to the outer joint and the outer joint to the end of the pendulum, respectively (Figure 4.1). The discrete Lagrangian is therefore

$$\begin{aligned} L_d \quad & (\theta_{1,k}, \theta_{1,k+1}, \theta_{2,k}, \theta_{2,k+1}) = && (4.57) \\ & \frac{1}{2}(m_1 + m_2) l_1^2 \left(\frac{\theta_{1,k+1} - \theta_{1,k}}{h}\right)^2 + \frac{1}{2} m_2 l_2^2 \left(\frac{\theta_{2,k+1} - \theta_{2,k}}{h}\right)^2 && (4.58) \\ & + m_2 l_1 l_2 \left(\frac{\theta_{1,k+1} - \theta_{1,k}}{h}\right)\left(\frac{\theta_{2,k+1} - \theta_{2,k}}{h}\right) \cos\left(\frac{\theta_{1,k+1} + \theta_{1,k} - \theta_{2,k+1} - \theta_{2,k}}{2}\right) + && (4.59) \\ & (m_1 + m_2) g l_1 \cos\left(\frac{\theta_{1,k+1} + \theta_{1,k}}{2}\right) + m_2 g l_2 \cos\left(\frac{\theta_{2,k+1} + \theta_{2,k}}{2}\right) && (4.60) \end{aligned}$$

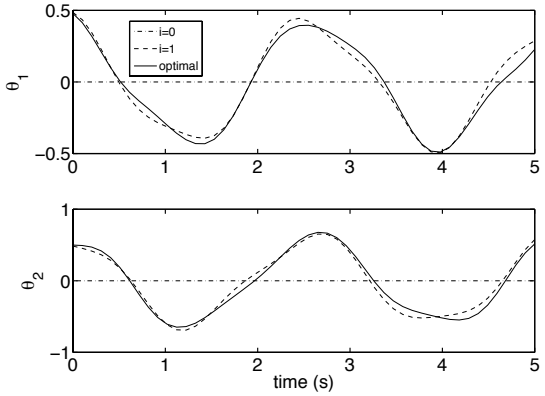

Figure 4.2: A comparison of the first two iterates of the projection method with the cost function $J(q_i, u_i) = \sum_{i=0}^{n-1} u_i^T u_i$ to the unforced simulation trajectory with initial conditions $\theta_1 = \theta_2 = \frac{1}{2}$.

We used the forced DEL equations and the cost function (4.53) to simulate the unforced system. Because the solution was known, it allowed us to test convergence characteristics, such as the number of iterations to reach the optimum, order of convergence, and accuracy of solution. The parameter values were $g = 9.81, m_1 = m_2 = l_1 = l_2 = 1$. For the first test, initial angle values were $\theta_{1,0} = \frac{1}{2}$ and $\theta_{2,0} = \frac{1}{2}$, with zero initial velocities. The time step h was set to 0.1 over a total of 50 time steps, or 5 seconds. The initial trajectory guess consisted of zeros for all variables.

Projection with Newton's method converged to the optimal (unforced) solution within 7 iterations, whereas it took numerical SQP 58 iterations to converge. Furthermore, the projection method approximated the exact solution very closely after just one iteration (Figure 4.2), whereas numerical SQP took significantly longer to do so. Projection with Newton's method reduces the error in the trajectory, as measured via the 2-norm, by 85% within one iteration (Figure 4.2), whereas numerical SQP reduces the error by only 10% (Figure 4.3).

For the second test, initial angle values were $\theta_{1,0} = \frac{\pi}{2}$ and $\theta_{2,0} = \frac{\pi}{2}$, with zero initial velocities. The time step h was set to 0.1 over a total of 50 time steps, or 5 seconds. All

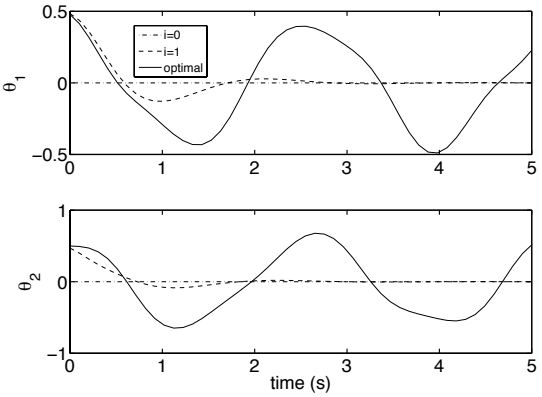

Figure 4.3: A comparison the first two iterates of numerical SQP with the cost function $J(q_i, u_i) = \sum_{i=0}^{n-1} u_i^T u_i$ to the unforced simulation trajectory with initial conditions $\theta_1 = \theta_2 = \frac{1}{2}$.

methods were given the same input, consisting of the optimal trajectory perturbed by value of $\epsilon = .001$ with control taking on an initial value of 10 for all time steps.

As can be seen in Figure 4.4, numerical SQP converged to a different solution than did the projection method because it was compromised by the control values. Furthermore, it took only 6 iterations for the projection method to converge to the correct trajectory and 428 iterations for numerical SQP to converge to a less than optimal trajectory (Figure 4.5). Gradient descent was allowed to run for 10^3 iterations, at which point it had still not yet converged, and is therefore not shown. Note that if given a more accurate controls trajectory, numerical SQP will converge to the optimal solution, but that it requires both trajectory and control variables to be reasonably correct to converge to the optimal value.

Figures 4.4 and 4.5 show a comparison of the 3 methods. As can be seen in Figure 4.5, in addition to converging to a less than optimal solution, numerical SQP converges more slowly. Projection with Newton's method converges quadratically, whereas numerical SQP converges, at best, superlinearly, and gradient descent converges only linearly. While it is likely that SQP's convergence could be improved if it were given more derivative information, it likely would not reach the convergence speed of projection with Newton due the the linearization of the constraints in SQP.

Due to its effectiveness at finding unforced trajectory with an unconstrained endpoint, it may be possible to use this method to solve the unforced DEL equations when they are degenerate. Initial results from simpler systems, such as the spring and the pendulum, show that, for short time horizons, projection with Newton's method is faster than root-finding.

4.5.2 Terminal Cost

To impose an endpoint on the system, the cost function was adjusted to include a steep penalty for deviating from the final condition. For an arbitrary system, this can be written

$$J_T = \sum_{i=0}^{n-1} u_i^T u_i + C_T(q_n - q_{end})^T(q_n - q_{end}), \tag{4.61}$$

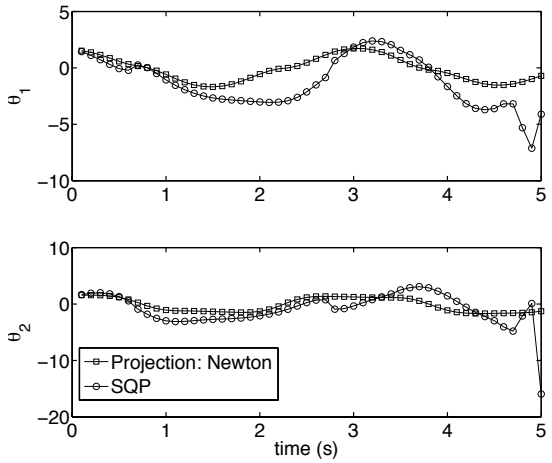

Figure 4.4: Solutions using numerical SQP and Newton's method with a cost of $J(q_i, u_i) = \sum_i^n u_i^T u_i$.

This DMOC problem yields a simulation of the system if it converges, but neither gradient descent (with ran for 1000 iterations without terminating) nor numerical SQP (which ran for 428 iterations to arrive at a non-optimal solution) implementation are able to converge. Newton's method converges without difficulty in 6 iterations to the actual optimal (unforced) solution.

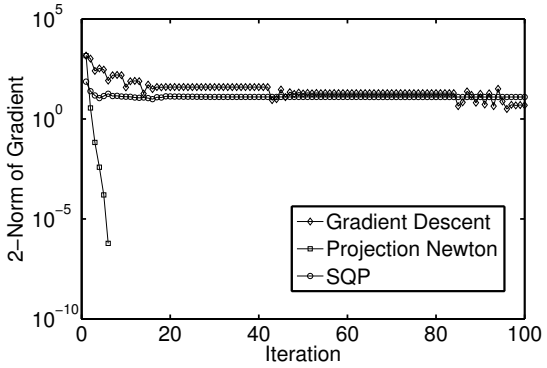

Figure 4.5: Log of the 2-norm of the gradient versus iteration for gradient descent, numerical SQP, and Newton's method with a cost of $J(q_i, u_i) = \sum_i^n u_i^T u_i$ for the first hundred iterations.

Gradient descent displays linear convergence, numerical SQP displays superlinear convergence, and Newton's method displays quadratic convergence, but neither gradient descent nor SQP has a high likelihood of converging to the correct optimal trajectory with a reasonable number of iterations.

where q_{end} indicates the required ending condition in configuration space and C_T is a coefficient representing the relative cost of not reaching the endpoint. For the double pendulum, this gives rise to the cost function

$$J_T = \sum_{i=0}^{n-1} u_i^T u_i + C_T[(\theta_{1,n} - \theta_{1,end})^2 + \tag{4.62}$$

$$(\theta_{2,n} - \theta_{2,end})^2 + \left(\frac{(\theta_{1,n} - \theta_{1,n-1})}{h} - \dot{\theta}_{1,end}\right)^2 + \tag{4.63}$$

$$\left(\frac{(\theta_{2,n} - \theta_{2,n-1})}{h} - \dot{\theta}_{2,end}\right)^2] \tag{4.64}$$

with $C_T = 1000$. C_T was set to be three orders of magnitude larger than the typical control values to assure the ending condition was met with reasonable accuracy. Increasing C_T beyond this value did not significantly affect results.

To test the effectiveness of convergence with terminal cost, we first compared gradient descent, numerical SQP, and the projection method with the initial condition $\theta_1 = \theta_2 = 0$, and the final condition $\theta_1 = \theta_2 = 0.7$. The time step value was set to $h = 0.1$ for 50 time steps, and all methods were given an input of initial trajectory and control of all zeros.

Projection with Newton's method converged to the optimal solution in 7 iterations, whereas numerical SQP took 47 iterations. However, tracking the trajectories for each iteration shows further advantages of using Newton's method. In one iteration, the difference between the optimal and estimated trajectories, measured via the 2-norm, is reduced by 92.5% (Figure 4.6), whereas numerical SQP reduces this value by only 7.5% (Figure 4.7). Past one iteration, the trajectories given by the projection method obscure the optimal trajectory.

In another test, we used the same three methods, gradient descent, numerical SQP and projection with Newton's method to invert the double pendulum, an unstable process. Both initial θ and $\dot{\theta}$ values were set to 0, with final θ values were set to π and $\dot{\theta}$ values again being 0. The time step was set to $h = 0.1$ for 40 time steps or 4 seconds.

All three methods were again given an input of an optimal trajectory perturbed by

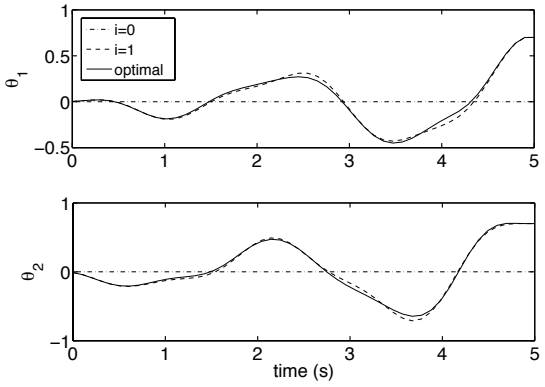

Figure 4.6: A comparison of the first two iterates of the projection method via Newton with the cost function $J_T(q_i, u_i) = \sum_{i=0}^{n-1} u_i^T u_i + C_T(q_n - q_{end})^T(q_n - q_{end})$ to the ultimate optimal simulation trajectory with initial conditions $\theta_1 = \theta_2 = 0$ and ending conditions $\theta_1 = \theta_2 = 0.7$.

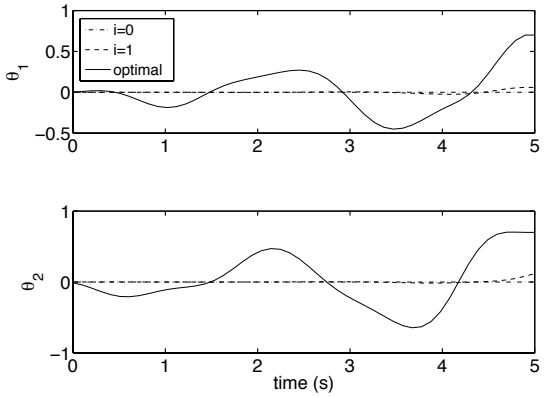

Figure 4.7: A comparison of the first two iterates of numerical SQP with the cost function $J_T(q_i, u_i) = \sum_{i=0}^{n-1} u_i^T u_i + C_T(q_n - q_{end})^T(q_n - q_{end})$ to the ultimate optimal simulation trajectory with initial conditions $\theta_1 = \theta_2 = 0$ and ending conditions $\theta_1 = \theta_2 = 0.7$.

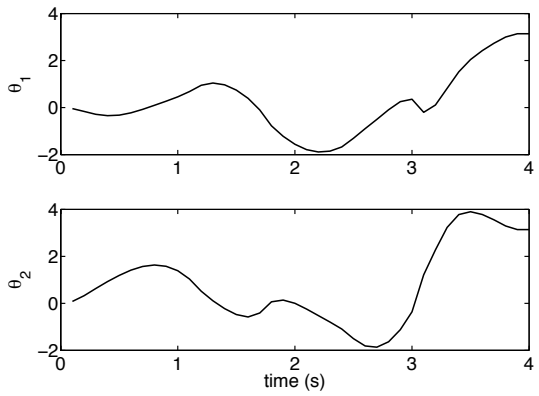

Figure 4.8: Optimal trajectory given by numerical SQP and projection with Newton's Method using the cost function $J_T \sum_{i=0}^{n-1} u_i^T u_i + C_T(q_n - q_{end})^T(q_n - q_{end})$.

The trajectory displays a pumping trajectory used to efficiently invert the double pendulum, taking it from initial angles of $\theta_1 = \theta_2 = 0$ to ending conditions $\theta_1 = \theta_2 = \pi$

ϵ=0.001. In this case however, rather than giving zero control, the control was also perturbed by 0.001. Both SQP and Newton's method via projection converged to the same optimal trajectory (Figure 4.8) with Newton converging in 4 iterations and numerical SQP taking 64. Gradient descent again failed to converge within 1000 iterations. One issue with gradient descent for this problem is that the step size for a given gradient direction is quite small, on the order of $10^{-4} - 10^{-5}$, slowing convergence.

A performance comparison of the three methods is shown in Figures 4.8 and 4.5.2. Clearly, projection converges much faster than numerical SQP or gradient descent. Furthermore, because projection depends only on the trajectory values, with the controls calculated as part of the projection, it can be more robust to perturbation. Specifically, if the trajectory is perturbed by as little as .001, and the controls entered as zero, projection will converge to the optimal trajectory, but numerical SQP will converge to a non-optimal solution.

4.5.3 Constrained Endpoint

Traditional DMOC uses a constrained endpoint, which creates the need to implicitly solve for the first free time step of the configuration variable in terms of the remaining

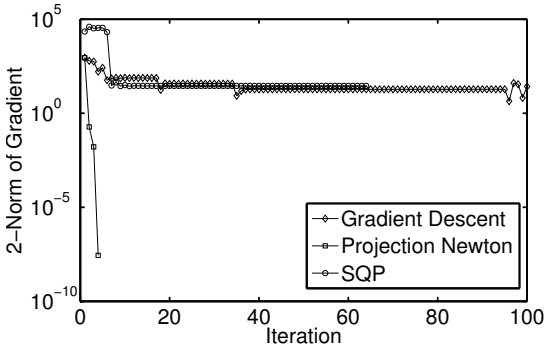

Figure 4.9: Log of the 2-norm of the gradient versus iteration for gradient descent, numerical SQP, and projection with Newton's method for the first hundred iterations with a cost of $J_T \sum_{i=0}^{n-1} u_i^T u_i + C_T(q_n - q_{end})^T(q_n - q_{end})$.

Both numerical SQP and projection with Newton's method converged to the optimal trajectory in 4 and 64 iterations, respectively, but gradient descent did not converge within 1000 iterations. Gradient descent exhibits linear convergence, numerical SQP superlinear convergence, and projection with Newton's method quadratic convergence.

configuration variables. This implicit differentiation can significantly complicate calculating the Hessian and sometimes lead to ill-conditioned matrices.

For a simple system, such as the single pendulum, the first configuration variable can be explicitly solved for, due to cancellation of terms in the DEL equations, which are

$$\begin{aligned} 0 &= u_1 + l\dot{\theta}_0 - \frac{l}{h}(\theta_1 - \theta_0) - \frac{gh}{2}\sin\left(\frac{(\theta_0 + \theta_1)}{2}\right) && (4.65)\\ 0 &= u_{i-1} + u_i - \frac{l}{h}(\theta_i - 2\theta_{i-1} + \theta_{i-2}) - && (4.66)\\ & \frac{gh}{2}\left(\sin(\frac{(\theta_i + \theta_{i-1})}{2}) + \sin(\frac{(\theta_{i-1} + \theta_{i-2})}{2})\right) && (4.67)\\ 0 &= u_n + l\dot{\theta}_n + \frac{l}{h}(\theta_n - \theta_{n-1}) - \frac{gh}{2}\sin\left(\frac{(\theta_n + \theta_{n-1})}{2}\right). && (4.68)\\ & && (4.69) \end{aligned}$$

where we again have the cost functional

$$J(q_i, u_i) = \sum_{i=0}^{n-1} u_i^T u_i. \tag{4.70}$$

A simulation was performed on this system for $l = 1$. Time steps of both $h = 0.1$ and $h = 0.05$ were used to test the effectiveness of projection as compared to numerical SQP with initial condition 0, ending condition $\frac{\pi}{2}$ for 101 time steps. An initial input of zeros for both control and trajectory was given for both size time steps. Both methods converge to a gradually pumping trajectory, but as can be seen in Figures 4.10 and 4.11, projection with Newton's method significantly outperforms both numerical SQP and gradient descent.

4.6 Preliminary Impact Work

4.6.1 Introducing an Impact

We have previously derived a method that allows us to, in DMOC, project onto the constraints and then minimize a given function using Newton's method. We wish to adjust this method in order to allow for an impact. Recall that we had previously found a projection

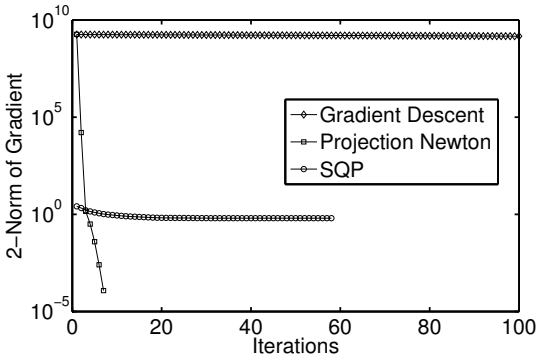

Figure 4.10: Log of the 2-norm of the gradient of the cost using gradient descent, numerical SQP, and projection with Newton's method for the single pendulum with an ending condition of $\frac{\pi}{2}$ with a time step of h=0.1.

Numerical SQP converges at 58 iterations, and projection with Newton converges after only 7. Gradient descent exhibits linear convergence, numerical SQP superlinear convergence, and projection with Newton's method quadratic convergence.

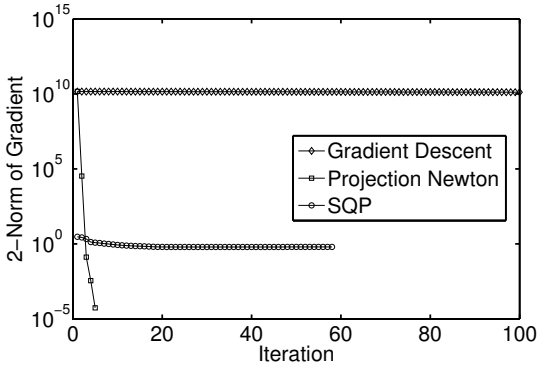

Figure 4.11: Log of the 2-norm of the gradient of the cost using gradient descent, numerical SQP, and projection with Newton's Method for the single pendulum with an ending condition $\frac{\pi}{2}$ and a time step of h=0.05.

Numerical SQP converges at 58 iterations, and projection with Newton converges at only 5. Gradient descent exhibits linear convergence, numerical SQP superlinear convergence, and projection with Newton's method quadratic convergence.

for the case without impacts. Up until an impact, we can use the same equations found previously to solve for the appropriate control variables

$$u_0 = -g_0(q(0), \dot{q}(0), q_0, q_1). \tag{4.71}$$

$$u_k = (-1)^i g_1(q(0), \dot{q}(0), q_0, q_1) + \sum_{j=1}^{k} (-1)^j g(q_{j+1}, q_j, q_{j-1}) \tag{4.72}$$

for $k = 1, \ldots n-1$.

We thus get the projection of the control variables onto the values that fulfill the constraints

$$\mathscr{P}_u(q, u) = (q, u(q)). \tag{4.73}$$

Including an impact introduces some complexity. Up until the impact, one projection is sufficient, as we have the same DEL equations. However, during and post impact, we need to define another projection so that we do not allow for configuration variables that go through the impact surface. Before the impact, there is no change in configuration variables, or in the projection onto allowable controls. However, after the impact, we need to define both the configuration and control variables based on the input variables.

Some additional notation is required to describe the equations around impact. Assume the impact occurs between time q_b and q_a. Denote the configuration space when impact occurs as q_{imp}, which can be solved for via the equation $\phi(q_{imp}) = 0$, where ϕ defines the impact surface.

We first need to project the configuration variables so that they bounce off the impact surface rather than going through it. We can thus allow for any given input of configuration variables, and then project the configuration variables orthogonally off of the constraint describing the impact surface, where the impact surface C is $q : \phi(q) = 0$. For example, in the case of a ball that bounces off the floor, $\{C = q : q = 0\}$, where q is the configuration variable representing the height of the ball. Thus, it is possible to have input heights of less than 0 (below the floor). We then define the projection to be

$$\begin{aligned} \mathscr{P}_q(q) &= q \qquad\qquad \phi(q) > 0 \\ \mathscr{P}_q(q) &= \mathscr{P}_{C,\perp}(q) \quad \phi(q) < 0 \end{aligned}$$

For example, for the bouncing ball, we would just have

$$\mathscr{P}_q(q) = \mid q \mid \tag{4.74}$$

This projection of the configuration variables must be done first before the projection involving the control variables is performed because we want to solve for for the controls using only allowable configuration variables. So for all of the following equations, it is assumed that we have already projected onto feasible post-impact configuration variables.

Without control, the equations involved in the impact are as follows

$$\begin{aligned} 0 &= D_2L_d(q_{b-1}, q_b, h) + D_1L_d(q_b, q_{imp}) \\ 0 &= D_3L_d(q_b, q_{imp}, \alpha h) - D_3L_d(q_{imp}, q_a, (1-\alpha)h) \\ 0 &= \lambda \bigtriangledown \phi(q_{imp}) + D_2L_d(q_b, q_{imp}, \alpha h) + D_2L_d(q_{imp}, q_a, (1-\alpha)h) \end{aligned}$$

Because the previous equations all involved the same time step, and because we only needed derivatives with respect to the configuration variables, we did not include this information in the equations. However, during an impact the time step varies, and completeness requires that we include this information in the relevant equations.

With control, we then have the equations, adjusted to

$$\begin{aligned} 0 &= D_2L_d(q_{b-1}, q_b, h) + D_1L_d(q_b, q_{imp}, \alpha h) + u_{b-1} + \alpha u_b \\ 0 &= D_3L_d(q_b, q_{imp}, \alpha h) - D_3L_d(q_{imp}, q_a, (1-\alpha)h) + \frac{2(q_a - q_{imp})u_b}{(1-\alpha)h} + \frac{2(q_{imp} - q_b)u_b}{\alpha h} \\ 0 &= \lambda \bigtriangledown \phi(q_{imp}) + D_2L_d(q_b, q_{imp}, \alpha h) + D_2L_d(q_{imp}, q_a, (1-\alpha)h) + u_b \end{aligned}$$

Because of the way we have defined the control variables for the non-impact case, the control terms become much more complex, particularly for the second equation.

Using these forms of the equations, we would have to solve these equations for λ, α, and u_b using an appropriate root finder. However, there are some problems with this method. There are multiple solutions. Therefore to get a feasible solution, the root-finder must have appropriate initial conditions and/or constraints on the values of the solutions. This leads to a smaller constrained optimization problem within the overall constrained optimization problem.

One way account for this complexity is to make some relatively minor adjustments to the problem that allow us to explicitly solve for the control variables. This allows us to form a similar projection as for the non-impact case in which the projection is quite simple and user-friendly. These simplifications are two-fold. First, we assume that the impact takes place exactly in the middle of the timestep, which means we always have $\alpha = \frac{1}{2}$. Secondly, we split the control value during the impact step in two, allowing there to be a different control before impact than after impact. This leads to the following equations

$$\begin{aligned}
0 &= D_2L_d(q_{b-1}, q_b, h) + D_1L_d(q_b, q_{imp}, \frac{1}{2}h) + u_{b-1} + \frac{1}{2}u_b \\
0 &= D_3L_d(q_b, q_{imp}, \frac{1}{2}h) - D_3L_d(q_{imp}, q_a, \frac{1}{2}h) + \frac{4(q_a - q_{imp})u_{b1}}{h} + \frac{4(q_{imp} - q_b)u_{b2}}{h} \\
0 &= \lambda \bigtriangledown \phi(q_{imp}) + D_2L_d(q_b, q_{imp}, h) + D_2L_d(q_{imp}, q_a, \frac{1}{2}h) + u_{b1} + u_{b2}
\end{aligned}$$

One major concern with this method is that the third equation becomes extraneous. If this represented a problem, it would be with the preservation of momentum. Though results thus far have not shown any problems due to this issue, it does represent a limitation and should be accounted for. While it should be resolved when the simplifications described previously are no longer used, it is worth noting here.

That caveat aside, with this simplification, we can merely solve for u_{b_1} and u_{b_2} fairly

simply via the first two equations. This gives

$$\begin{aligned} u_{b_1} &= 2\left(-D_2L_d(q_{b-1},q_b,h) - D_1L_d(q_b,q_{imp},\frac{1}{2}h) - u_{b-1}\right) \\ u_{b_2} &= \frac{\frac{1}{2}h}{2(q_{imp}-q_b)}\left(-D_3L_d(q_b,q_{imp},\frac{1}{2}h) + D_3L_d(q_{imp},q_a,\frac{1}{2}h) - \frac{2(q_a-q_{imp})u_{b1}}{\frac{1}{2}h}\right). \end{aligned}$$

Once we solve for u_{b_2}, we can proceed to the other controls for post-impact. We start with the last remaining equation involving any of the impact variables,

$$0 = D_2L_d(q_{imp},q_a,(1-)h) + D_1L_d(z_a,z_{a+1}) + \frac{1}{2}u_{b_2} + u_a, \tag{4.75}$$

which can be solved as

$$u_a = -D_2L_d(q_{imp},q_a,\frac{1}{2}h) - D_1L_d(z_a,z_{a+1}) - \frac{1}{2}u_{b_2}. \tag{4.76}$$

We then have the feedback law from above take affect again for all $k > a$

$$u_k = -u_{k-1} - g(q_{k-1},q_k,q_{k+1}). \tag{4.77}$$

This feedback law will continue to hold until another impact occurs, at which point, the impact equations are again introduced and solved. The simplified projection does allow for multiple impacts to occur, though, as will be seen in the following section.

So for these variable, we have a combination of proejctions, defined by

$$\mathscr{P} = \mathscr{P}_u(\mathscr{P}_q(q,u)). \tag{4.78}$$

Though when we do not simplify in order to find the controls around the impact, it may change the projection slightly, we will always have to project first onto feasible configuration variables and then find a projection that allows for appropriate controls.

Once, the projection has been found, we need to find the appropriate derivatives in order to perform the optimization. These derivatives can again be found using

$$DJ_c = DJ(\mathscr{P}(q)) \cdot D\mathscr{P}(q), \tag{4.79}$$

where J denotes the unconstrained cost function and J_c the cost function constrained via the projection. Similarly, the Hessian can be calculated via

$$\begin{aligned} D^2 J_c &= D^2 J(\mathscr{P}(q)) \cdot D\mathscr{P}(q) + \\ & \quad DJ(\mathscr{P}(q)) \cdot D^2 \mathscr{P}(q) \end{aligned}$$

where the elements of $D\mathscr{P}$ and $D^2\mathscr{P}$ can be found by taking the first and second derivatives the projection elements as defined via the DEL equations.

Once we have found our projection, it remains to calculate the derivatives. The complexity of the derivatives depends upon both the system, which determines P_u, and the impact surface, which can make P_q's derivatives quite involved even for a very simple system. There can be some complications around the impact surface, but they are generally minor and avoidable by slightly changing the length of the time step. These problems generally seem to stem from getting within numerical precision of the impact surface, but this postulate has not been rigorously proven as of yet.

We originally attempted to find optimal trajectory and control variables using Newton's method, but found that the basin of attraction for Newton was too small for it to be effective. Instead, we chose to employ a hybrid approach, using trust region methods when the Hessian (calculated from the actual second derivatives, not using finite difference) is not positive definite, but switching to Newton when it is. This combines the large basins of attraction that are possible with trust region methods with the fast convergence seen by Newton. Initial exploration for the following problems has indicated that in terms of the ending trajectory, it rarely if ever makes a difference whether trust region or Newton's method is used, but that Newton converges faster.

With impacts, however, additional complexity is introduced into the system because the control at the impact step depends on all configuration variables for the time steps previous to impact, whereas oftentimes, without an impact, control values depend only on the most immediate configuration variables. This additional dependence leads to a much less

sparse Hessian than occurs in systems without impacts. This complexity is increased with each additional impact. Though this does not, seemingly, significantly affect the projection method, it is possible that it is problematic for SQP, which requires a sparse Hessian for good convergence.

Another important aspect of the implementation is that it must be able to correctly identify when the impact should occur in the trajectory for every iteration, and even within an iteration. For this reason, we need to use the input trajectory, the trajectory that goes through the impact surface, throughout the process. If we do not allow the input trajectory to go through the surface, the impact point cannot be correctly identified and must therefore remain constant, which limits the efficacy of the algorithm. Because the overall projection acts on this input trajectory, we also base the search direction on this trajectory and continuously update it throughout the optimization process.

4.6.2 Bouncing Ball: 1 Impact

No Endpoint We have implemented this combination of projections for a very simple system, a bouncing ball, which starts from a given height, and then falls. If it reaches a value of 0, it encounters an impact surface, off which it bounces, perfectly elastically. The cost function for the ball without an endpoint was

$$J(q_i, u_i) = \frac{1}{2}\sum_{i=0}^{n-1} u_i^T u_i \tag{4.80}$$

We first tested to see if the method would converge. Given an initial height of 2 meters and a timestep of 0.05 seconds, an initial input trajectory of all ones led to quite fast convergence (Figure 4.12). We also see that the impact can change time steps when it incurs a significant decrease in cost (Figure 4.12).

Given this initial condition, using the projection method, the impact time changed twice (Figure 4.13). Because the initial input trajectory was entirely positive, no impacts

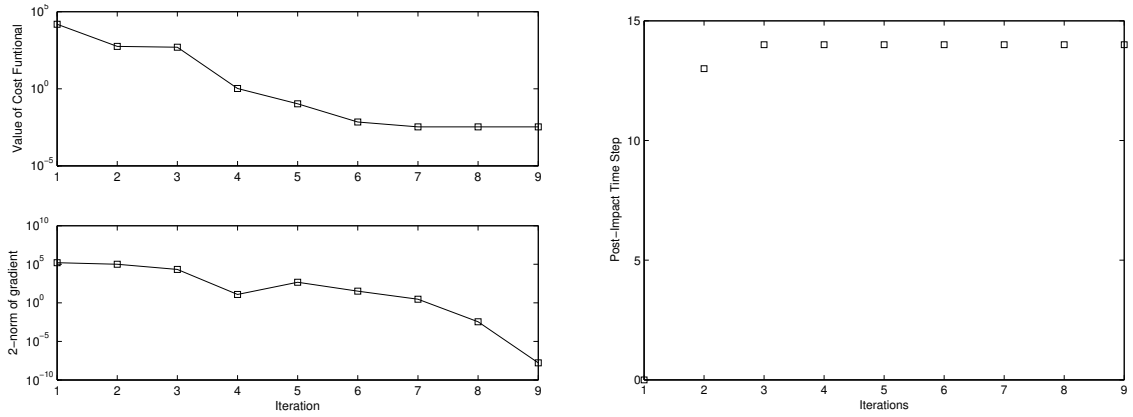

Figure 4.12: Projection converges quickly for an input of all ones. SQP stalls, and is unable to converge at all for this input condition.

occurred during this first iteration. However, for this particular problem, assuming no impact, the system converges to the optimal solution in one step. Therefore the second iteration was the exact trajectory that would occur if there were no impact surface. This trajectory went from positive to negative between time steps 12 and 13. The impact time switched again during the next step from between steps 12 and 13 to between steps 13 and 14. Because the projection is defined based upon when the impact occurs, which affects the search direction, we cannot expect it to always change, but a significant change in cost due to an impact time change will allow a change to sometimes occur.

We tried the same input trajectory with SQP, and the algorithm did not converge because the updates to the Hessian led it to become ill-conditioned. Furthermore, it took very few iterations before the algorithm stalled.

However, this is not a universal problem, rather it depends upon the initial condition given to SQP. We altered the initial conditions so that the first half of the input trajectory was still ones, but the second half was negative ones, with all other parameters remaining the same. This change did allow SQP to converge, although, as can be seen in the following figure, not as quickly as could the projection method (Figure 4.14).

Though the two methods converge to the same trajectory here, because we are using a gradient descent based algorithm, it is not the global optimum. In fact the trajectory in the first example is not optimal either. In the case of the first trajectory, without the simplification, the impact would occur quite close to one of the time steps, rather than exactly between 2. It is possible that this problem is pushing the impact into the next time step.

One of the issues for projection is that it tends to move the step during which the impact occurs forward in time, even if it might be advantageous to move it backward in time. It is possible that this is due to the restriction that the impact occurs halfway through the step. This may make a little "bump" which may be compromising the optimization process. However, we cannot be certain that this is what is going on or if there is a bigger

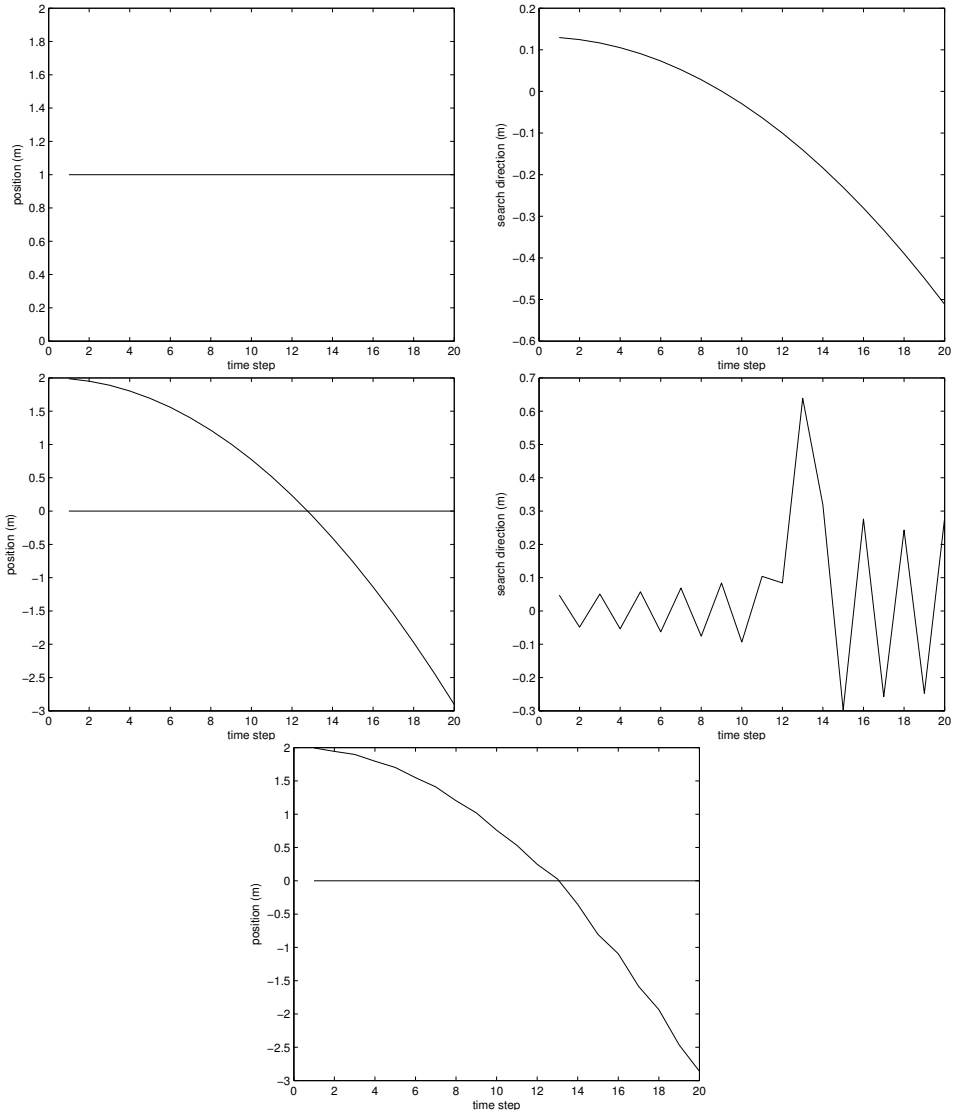

Figure 4.13: It is possible for an impact point to change, but because the projection is based on the impact, it is a rare occurrence.

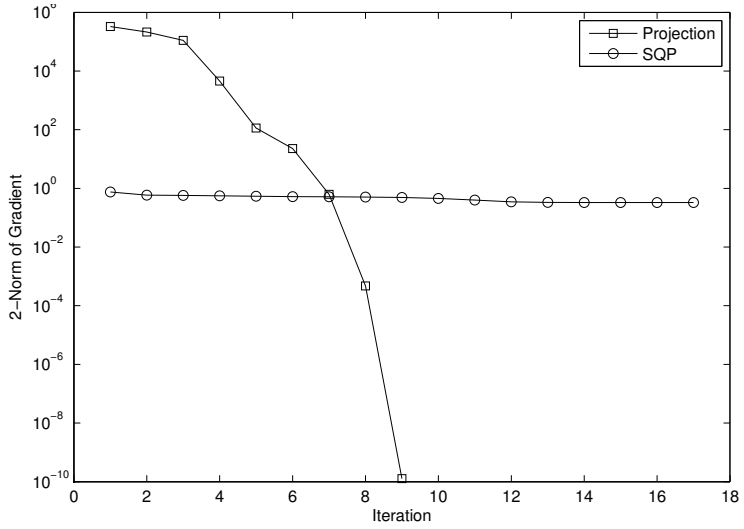

Figure 4.14: Projection does converge faster than SQP for the cases where SQP converges.

problem. We will do more exploration of this asymmetric impact step change when we remove the simplification requiring the impact to take place halfway through a step.

Adding an Endpoint We can add an endpoint to this problem by modifying the cost function to add a prohibitive cost of deviating from the endpoint. We impose an ending condition of $q_n = 1m, \dot{q}_n = 0$ via the cost function

$$J(q_i, u_i) = \frac{1}{2}\sum_{i=0}^{n-1} u_i^T u_i + C_T(q_n - q_e)^T(q_n - q_e) \tag{4.81}$$

where C_T in this case is 100. In terms of implementation, because we use finite differences to get the derivatives, this required a little modification to make sure one cost did not dominate. We modified the value for the coefficient for the velocity term so that neither term dominated.

We still get better convergence for projection (Figure 4.15), but again it's not to the ultimate optimum, at least for the initial conditions we've chosen. We did find that projection performs better than does SQP, as it converges to a more optimal trajectory by $\sim 15\%$, which is a definite improvement (Figure 4.15). However, the actual absolute optimimum for these parameters is somewhere between the two and has a cost approximately half of what is found

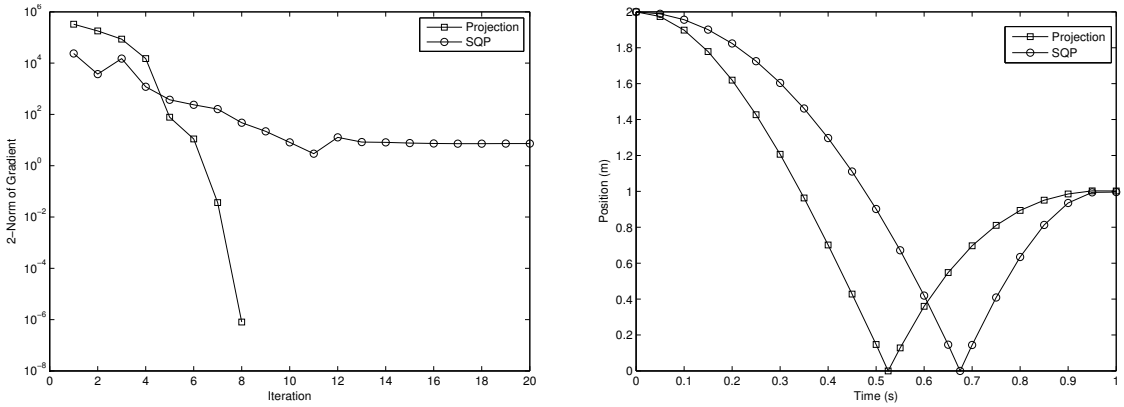

Figure 4.15: A comparison of SQP and projection with an endpoint included.

with projection. Projection impacts the surface too early, and SQP impacts the surface too late. So, though projection is an improvement on SQP, because of its construction, it does not reach the global optimum.

4.6.3 Bouncing Ball : 2 Impacts

No Endpoint To introduce another impact to the same system described previously, we used the same parameters that were used for the previous system, but added time steps. We doubled the length of the trajectory, and used an input trajectory consisting of all ones,

We see the same essential advantages and disadvantages of projection over SQP with two impacts as we did with one impact, with one difference (Figure 4.16). We have yet to get SQP to converge at all. Even for one impact, if we fix the impact step, that is, restrict the impact to occurring between two specific time steps, SQP will work. However, we have been unable to find any combination of parameters and initial inputs that will allow SQP to converge with two impacts. It runs into the same problem with an ill-conditioned Hessian as occurred with a single impact.

We rarely see changes in the impact step for two impacts. For instance, if we give an input trajectory, where the first 10 time steps are ones, the next 20 are negative ones,

and the last 10 are ones, the impact steps do not change at all, even though it would be advantageous to do so (Figure 4.16).

One way we can deal with this, is to generally find where the impact might occur, and then give input trajectory values that restrict the trajectory to be close to the impact surface. Sometimes this will allow for more flexibility in the impact time step. We have had some success with using an input trajectory that has a pattern of linearly decreasing from the initial condition (at t=0) to very small values right around the impact, and then linearly increasing and decreasing until one might expect a second impact, where we again use very small values. However, this is somewhat suspect as it requires an *a priori* prediction of a likely trajectory in order to find an appropriate initial condition.

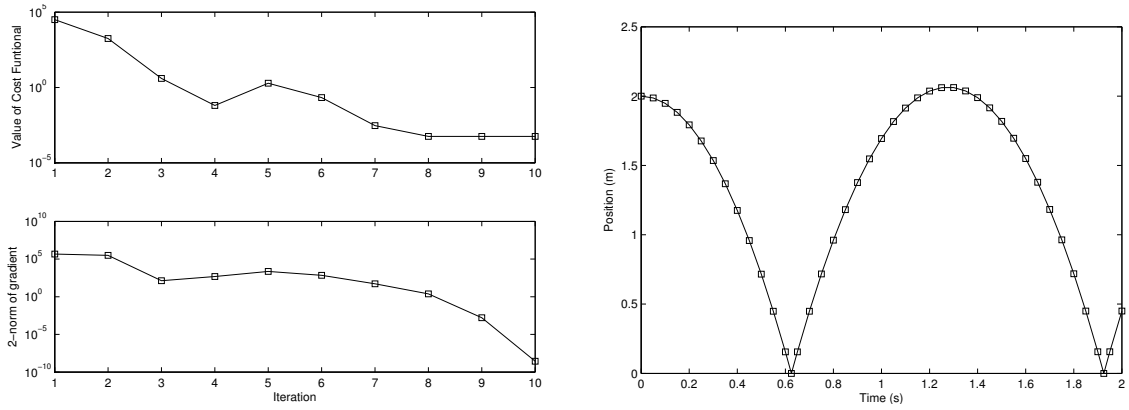

Figure 4.16: We see a very accurate trajectory and fast convergence with two impacts included.

Endpoint

We have done very limited work for an endpoint for 2 impacts with a constrained endpoint as well. We again used a terminal cost to constrain the endpoint, using the cost function described for one impact.

We get the same convergence pattern here as for the condition without an endpoint (Figure 4.17), but also the same problems. The impact step does not change unless given a specific type of trajectory (Figure 4.17). Furthermore, in the case where there is no

constrained endpoint, we can generally get very close to the optimal trajectory by using an input trajectory of all ones. Not having an impact on the first step leads to a trajectory that goes through the impact surface at an appropriate time, which is then projected, and the process is then repeated for the second impact. This gives at least one initial condition that always leads to fairly accurate results without an endpoint. However, this strategy does not work when there is a constrained endpoint. If we enter a trajectory of all ones when the endpoint is constrained, an impact is never introduced. We just get a trajectory that starts at the initial point and ends at the endpoint, and never encounters the impact surface at all (Figure 4.18).

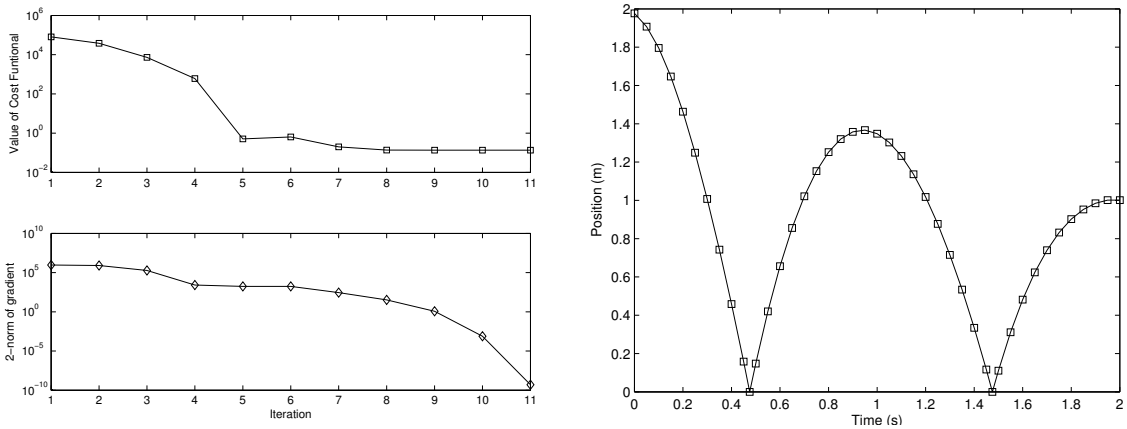

Figure 4.17: Projection still functions when there are multiple impacts, but SQP stalls.

Though the examples given are for short trajectories, we find the same patterns with longer trajectories. However, these general patterns are easier to see on the shorter trajectories than the longer ones.

4.7 Conclusions and Future Work

We present an approach to DMOC that provides local quadratic convergence that in practice is considerably faster than "black box" optimization techniques that utilize numerical differentiation approximations However, the current formulation is local, assumes full actuation and compares to SQP with numerical derivatives. Additional work will focus

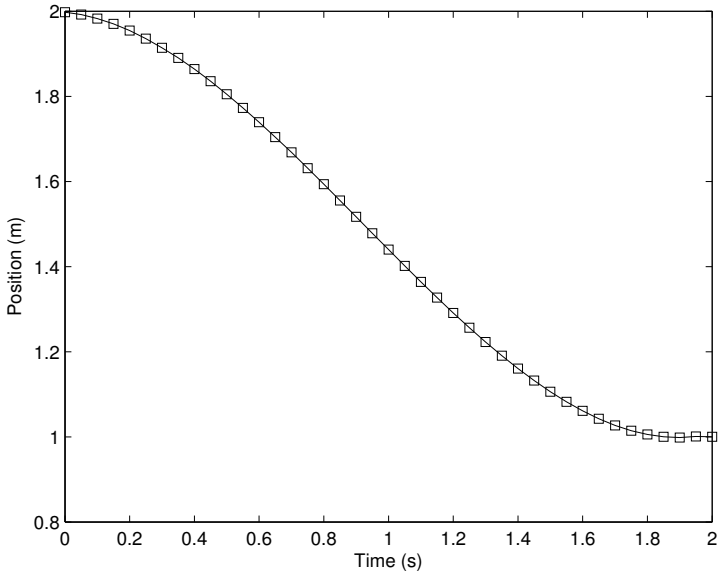

Figure 4.18: If the initial trajectory does not go through the impact surface, projection will never lead to an impact occurring.

on using implicit differentiation to differentiate the projection associated with constrained endpoints, which will allow for the more accurate simulation of an ending condition. Furthermore, all the systems discussed here have been fully actuated, whereas one of the advantages of DMOC is its ability to stably handle underactuated systems, so a next step will be adjusting the projection to be able to work with these types of systems. Lastly, to fully gauge the advantages of this method, we would need to input derivatives into SQP rather than calculating numerically calculating them. It is likely that initially using SQP to allow for a large basin of attraction and then projection when near the optimum could considerably improve convergence time while retaining a large basin of attraction, making a much more effective constrained optimization algorithm.

Chapter 5

Modeling Uphill and Downhill Running

Collaborators: Dr. Jinger Gottschall, Penn State University, Jason Franz, M.S., University of Colorado

5.1 Introduction

In forms of locomotion such as swimming or bicycling, the mechanical power output of the body predominantly determines the metabolic demand (Di Prampero et al., 1979; Capelli et al., 1993; Chavarren and Calbet, 1999). However, one of the fundamental disconnects in legged locomotion (i.e. walking and running) research is that the metabolic demand does not necessarily correlate to the mechanical power output (Cavagna et al., 1963). This disconnect is due to energy saving mechanisms, such as the ability of tendons to store and return elastic energy during running gaits. Because elastic energy is storage and return is both difficult to measure and estimate, mathematically modeling offers a way to explore the disconnect between mechanical power output and metabolic demand.

One reason for the disconnect between mechanical power output and metabolic demand is that skeletal muscle has three distinct types of behavior, each associated with a different cost. The first is concentric action, in which the muscle is active and shortens. This type of muscle action is used when standing up from a squat. The quadriceps muscles (front of thigh) contract concentrically to extend the knee and lift the body upward. The knee torque generated by the quadriceps is greater than the torque created by gravity, and thus the muscle shortens as the knees extend, performing mechanical work. Concentric muscle

actions are the most energetically expensive. A second type of muscle behavior is isometric action. In this case, the muscle is active, but remains at the same length. Standing stationary with the knees bent requires isometric action in order to keep the knees from collapsing. In this case, the torque generated by the muscle is equal to the torque generated by gravity, so no movement occurs. This action is less energetically expensive than concentric action, and ideally, no external mechanical work is performed by the muscle. A third type of behavior occurs when the muscles are active but lengthened by an external force. That is known as eccentric action. Going from standing to a squat requires eccentric action of the quadriceps. The muscles produce force that generates a torque, but it is less than the torque produced by gravity on the body. Consequently, the body is lowered in a controlled manner. Eccentric actions are 4-5 times less energetically energetically expensive than concentric contractions and 20% less expensive than isometric contraction. Tendons, unlike muscles, incur negligible metabolic cost. They are made of passive collagen, an elastic material capable of storing and returning elastic energy.

These ideas are particularly relevant during the stance phase of running. The stance phase of running is the period of time when the foot is actually on the ground, from the time of heel-strike to toe-off. During the first half of stance, as the ankle, knee, and hip joints flex, the center of mass (CoM) moves downward, and its velocity decreases. During the second half of stance, as the ankle, knee, and hip joints are extended, the CoM moves upward, and its velocity increases. It was originally thought that the extensor muscles act eccentrically during the first half of stance and concentrically during the second half, with all the length change occurring in the muscles and none in the tendons. This idea, however, was proven untrue when it was found that these stretch-shorten actions occur with an efficiency of much greater than the 25% *in vitro* (isolated outside the body) measurement of muscle efficiency (Cavagna et al., 1963).

It has therefore been determined that a substantial length change occurs in the Achilles tendon, and that a portion of the mechanical energy decrease of the CoM during the first

half of stance is stored in the Achilles tendon and then returned during the second half of stance. Because it is very difficult to measure the forces in the Achilles tendon *in vivo* (in a living animal), models have been built to approximate the leg's behavior.

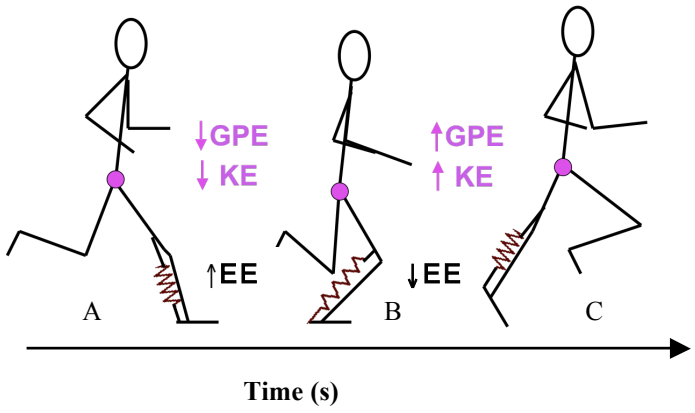

Figure 5.1: Relating human mechanics to the spring-mass model

The most basic model of running mechanics, the spring-mass model, simulates level running very well. In ideal level running, it is possible to have perfect elastic return because the energy of the CoM at the end of stance is equal to the energy of the CoM at the beginning of stance. A model has been built representing the leg as a compressive spring, with the mass of the body (m) concentrated at a point atop the spring. The spring-mass system is modeled by simultaneously solving the following differential equations for the position relative to the foot (d) and velocity (v) in the vertical (z) and fore-aft (y) directions from the time the foot strikes the ground until it leaves the ground. Note that the value d_y is negative if the mass is behind the foot.

$$\begin{aligned} \dot{d}_z &= v_z \\ \dot{v}_z &= -g + \frac{k}{m}(l_o - \sqrt{d_z^2 + d_y^2})\frac{d_z}{\sqrt{d_z^2 + d_y^2}} \\ \dot{d}_y &= v_y \end{aligned}$$

$$\dot{v_y} = \frac{k}{m}(l_o - \sqrt{d_z^2 + d_y^2})\frac{d_z}{\sqrt{d_y^2 + d_y^2}}$$

Though this is the typical system of equations for use in comparing simulations to experimental data, it is conceptually much easier to think of the model in polar rather than Cartesian coordinates, which leads to the following system of equations.

$$\begin{aligned}
\dot{r} &= v_r \\
\dot{v}_r &= \frac{k}{m}(l_o - r) - g\cos\theta + r\dot{\theta}^2 \\
\dot{\theta} &= v_\theta \\
\dot{v}_\theta &= \frac{g\sin\theta}{r} - \frac{2\dot{r}\dot{\theta}}{r}
\end{aligned}$$

where r and θ are defined in the following figure.

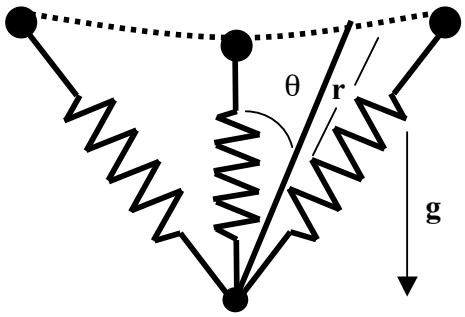

Figure 5.2: Spring-mass model polar coordinate variable definitions

The spring-mass model has been sufficient to model running at different stride frequencies (Farley and Gonzalez, 1996) and speeds (McMahon and Cheng, 1990; He et al., 1991) by varying the spring stiffness of the leg (k), and the touchdown angle, defined as the angle between the vertical directly above the foot and the leg at touchdown. However, the passive spring-mass model cannot represent running on slopes because it is impossible for elastic energy to sufficiently explain the net CoM movement in uphill and downhill running

(Gottschall and Kram, 2005). In uphill running, the potential energy of the CoM is greater at toe-off than heel strike, necessitating that some positive work (actuation in the form of concentric contraction) is performed by the muscles. The need for energy/work input is the reason why it is so much more difficult to run up even a shallow grade than to run on the level. In downhill running, the potential energy of the COM is less at toe-off than heel-strike, necessitating that some negative work (damping in the form of eccentric contraction) is performed by the muscles. The need to dissipate energy leads to the muscle soreness felt after a prolonged downhill run. Similar patterns are found in the kinetic energy of the COM with slope, though with smaller amplitude changes. This change in energy is inconsistent with the above system of equations because no energy input or output is allowed; it is an energy-conservative system.

However, there are ways to simulate mechanical energy change in a mathematical model. On the uphill, adding an actuator in series with the spring would represent the work done by the calf, thigh, and gluteus muscles for uphill running. For downhill running, adding negative actuation in series with the spring would represent the negative work done by the quadriceps muscles to dissipate the extra energy accumulated by running downhill at a steady speed. We hypothesized that using a simple actuation adjustment to the original spring-mass model could allow accurate modeling of uphill and downhill running, respectively. We further hypothesized that, for shallower slopes, the body adjusts touchdown angle in order to reduce actuation, but that, as the slope becomes steeper, this method becomes infeasible, forcing more actuation to occur.

5.2 Methods

The model represents a biological system, which is capable of processing and integrating feedback and adapting the motion accordingly. Consequently, a controls system perspective is a logical framework to use to analyze the problem. A controls framework allows feedback, based on measures that the body and brain can perceive, such as force. We used a linear

quadratic regulator model, which requires a reference trajectory. A reference trajectory can be constructed in a number of ways. The most accurate and easiest way to do so is to take experimental data from subjects.

5.2.1 Experimental Data

9 subjects (mass = 64.7 ± 11.8 kg , leg length = 0.94 ± 0.06 m) participated in this study. Before any data collection took place, subjects completed a short questionnaire to determine whether they were healthy and fit enough to complete the study All subjects gave informed consent before participation, and the protocol was approved by the University of Colorado Human Research Committee.

Subjects completed 4 separate experimental sessions. Due to the time intensive process of sloping the treadmill, these sessions were not randomized but were in the following order: level, 9°, 6°, and 3°. Before any data were collected, subjects walked for 5 minutes at 1.25m/s to familiarize themselves with the treadmill. Following this familiarization period, we placed 15 reflective markers on the lower limbs. We used an 8 camera system (Motion Analysis, Santa Rosa, CA) to collect 3-D kinematics at 100 Hz and a force plate (AMTI, Watertown, MA) instrumented treadmill to collect force data at 1000 Hz. For each session, subjects performed a static standing trial before any running trials to establish body weight and allow for marker calibration. All running trials were 30 seconds in length and were recorded while subjects ran at 3 m/s on the force-treadmill.

For the level session, subjects first ran for 30 seconds without any instruction during which we determined each subject's preferred stride frequency (PSF). The subjects then completed 3 trials in random order during which they ran at preferred stride frequency and 15% faster and slower than preferred stride frequency, as enforced by a metronome.

For the uphill and downhill trials, subjects performed 6 trials: uphill and downhill at preferred stride frequency and 15% above and below preferred. The subjects first ran for 30 seconds without any instruction uphill and downhill in random order. During these 30

seconds, their preferred stride frequency was determined. The subjects then completed 6 trials in random order during which a metronome enforced stride frequencies of preferred, 15% faster than preferred, and 15% slower than preferred while they ran uphill and downhill.

For all trials, we calculated the vertical and horizontal trajectories of the COM for each step as per Cavagna (1975) using a custom-written Matlab program. The program first filtered all force data using a 4th order Butterworth filter at 20 Hz. It then integrated horizontal and vertical ground reaction force data once to calculate instantaneous velocity in the horizontal (v_h) and vertical (v_v) directions. The velocity data were then integrated again to obtain instantaneous displacement relative to touchdown using zero as the integration constant. We used these data to calculate changes in the location of the CoM location over the course of stance phase. An average trajectory was then found by normalizing stance phase for each step to the average contact time and then taking the average over all steps. We additionally calculated the spring constant and touchdown angles for the model using

$$\theta_{td} = \sin^{-1}\left(dy_1/l_o\right) \tag{5.1}$$

$$k = \frac{F_{z_{max}}}{\Delta L} \tag{5.2}$$

where dy_1 is the distance the center of mass travels during the first half of stance, l_o is the leg length, $F_{z_{max}}$ is the maximum force in the vertical direction, and ΔL is the total change in leg length during the first half of stance as defined in (Farley and Gonzalez, 1996).

Because of changes in the location of the point of force application on the treadmill surface over the course of stance phase, the touchdown angle that is appropriate for the model does not generally exactly match the touchdown angle calculated from the angle between the greater trochanter (hip) and the medial malleolus (ankle). Due to this discrepancy, it was necessary to do a certain amount of approximation in calculating the touchdown angles appropriate for the model in uphill and downhill running. To estimate these angles, we took the difference between the experimental angles calculated for the level and hill angles. We then found the touchdown angle that led to the typical spring-mass dynamics on the

level based on the experimental data. We then adjusted the model level touchdown angle by the differences found in touchdown angles from the level to the various slopes to have a reasonable input value.

Once these analyses had been performed on all trials, we split the data into two groups randomly by subject. The first five subjects were used to identify the actuation necessary for the conditions studied. The remaining subjects were used to validate the model.

5.2.2 Model Construction

We first constructed a fully actuated system (Figure 5.2.2), one with control variables for both the leg length (r) and leg angle (θ), leading to the equations

$$\dot{r} = \dot{r} \tag{5.3}$$

$$\ddot{r} = \frac{k}{m}(l_o - r + a_1(t)) - g\cos\theta + r\dot{\theta}^2 \tag{5.4}$$

$$\dot{\theta} = \dot{\theta} \tag{5.5}$$

$$\ddot{\theta} = \frac{g\sin\theta}{r} - \frac{2\dot{r}\dot{\theta}}{r} + \frac{k}{m}a_2(t) \tag{5.6}$$

where $(a_1(t), a_2(t))^T = a(t) = -k_g(t)(x(t) - x_{ref}(t))$. The coefficient for the angle actuation was used to keep the relative values of length and angle actuation similar, but values reported in the results will be in rad/s.

Once the specific reference trajectory is found, the system must first be put in the necessary linear, time-dependent form in order to use the Riccati equation. In short, system must be of the form

$$\dot{x} = A(t)x(t) + B(t)u(t) \tag{5.7}$$

where $x(t)$ is the state vector, in this case leg length and angle position and velocities, and $u(t)$ is the control vector, in this case the actuation values. This can easily be done for our system for each phase.

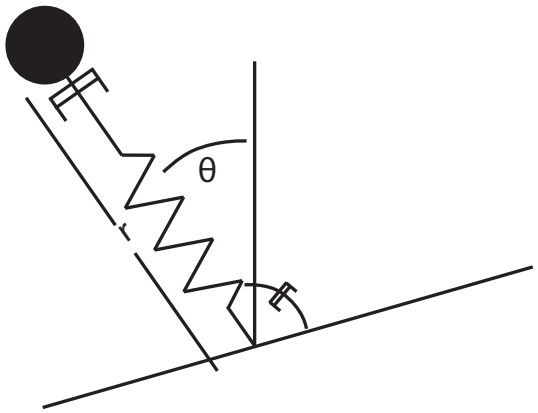

Figure 5.3: The fully actuated spring mass model.

Note that because the system is nonlinear, we have to approximate $A_s(t)$, which we do by taking the Jacobian of the system. Doing so gives the following matrix

$$A(t) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{k}{m} + \dot{\theta}^2 & 0 & g\sin\theta & 2\dot{\theta}r \\ 0 & 0 & 0 & 1 \\ \frac{g\sin\theta - 2\dot{r}\dot{\theta}}{r^2} & -\frac{2\dot{\theta}}{r} & -\frac{g\cos\theta}{r} & -\frac{2\dot{r}}{r} \end{pmatrix}$$

.

The system is also nonlinear and time-dependent in the control variables, so $B(t)$ again requires a Jacobian.

$$B_s(t) = \begin{pmatrix} 0 & 0 \\ \frac{k}{m} & 0 \\ 0 & 0 \\ 0 & \frac{k}{m} \end{pmatrix}$$

.

Having established the matrices $A(t)$ and $B(t)$ for each phase of a step, we wish to solve for the appropriate gains using

$$k_g(t) = R^{-1}B(t)P(t) \tag{5.8}$$

.

where $P(t)$ is found by solving following equation backward in time

$$\dot{P}(t) + P(t)A(t) + A(t)^T P(t) - P(t)B(t)R^{-1}B(t)^T P(t) + Q = 0 \qquad P(T) = P_1 \tag{5.9}$$

.

However, note that we have not yet defined the matrices Q and R. Before defining Q and R, it is important to establish what information they hold. The matrices Q and R come from the cost function which is being minimized. The typical form is

$$J = \int_{t_0}^{t} x(t)Qx(t) + x(t)Ru(t) \tag{5.10}$$

but in our case, we care about distance from a desired trajectory, giving

$$J = \int_{t_0}^{t} (x_{ref}(t) - x(t))Q(x_{ref}(t) - x(t)) + u(t)Ru(t) \tag{5.11}$$

Thus, Q, which is diagonal, represents how heavily the cost function depends on the difference between the actual and ideal trajectories only. Essentially, Q measures how much it costs for the actual trajectory to deviate from the ideal trajectory. Q is not found mathematically but determined by the relative cost of deviation of each dimension of the state space. For this problem, we have equated made the cost of deviating from each element of the state space $(r, \dot{r}, \theta, \dot{\theta})$ making $Q = I(4)$. Similarly, R is defined to be $I(2)$.

For the identification portion of the analysis, we found the actuation necessary for the 5 randomly chosen "identification" subjects. In order to determine how much of the actuation was due to the inherent difference between the subject and the model, we first found the actuation necessary on the level. We then found how much actuation was necessary to approximately match the experimental trajectory for each slope. We used the same values for the leg stiffness k for all slopes, but adjusted the touchdown angle based on the experimental data. To separate the effect of the slope on the actuation from that of the

subject, we subtracted the actuation on the level for a given subject from the actuation necessary for that subject on all of the slopes. To find the average actuation needed for a given slope, regardless of subject, we normalized the length actuation to contact length (the distance the center of mass moves horizontally during stance phase) and the angle actuation to initial angular velocity, and the stance time to percentage of totals stance time. Once these values were normalized, we averaged them for each slope over all subjects to find the typical non-dimensional actuation necessary.

Once we had this average actuation, we used it to validate the model. For the validation subjects, we took the average non-dimensionalized actuation, and renormalized it for each subject. To do this, we multiplied length actuation by contact length and angular actuation by initial angular velocity. We then added on the actuation for the given subject for the level. This pattern of actuation was used as an input for the model for validation purposes.

5.3 Results

Overall the model proved to be very effective, especially uphill (Figure 5.4). However, the model had a tendency to overestimate the necessary actuation across slope.

As can be seen from the identification data, the model is capable of accurately approximating the actuation necessary for running downhill (Figure 5.5). We do find larger differences between the experimental data and the actuated CoM trajectory with an increase in slope magnitude. These differences could be corrected for with a reduction in the magnitude of the elements of R, which effectively decreases the cost of actuation, but this adjustment leads to instability of the actuation values.

This discrepancy between identification and validation data led us to further explore the average actuation patterns and their variability for each slope. Figure 5.6 shows average actuation patterns for all subjects on all slopes on a single graph. Both the difference between the actuation on each slope and on the level and the raw actuation data are shown.

As shown in Figure 5.6, the length actuation shows a consistent pattern regardless of

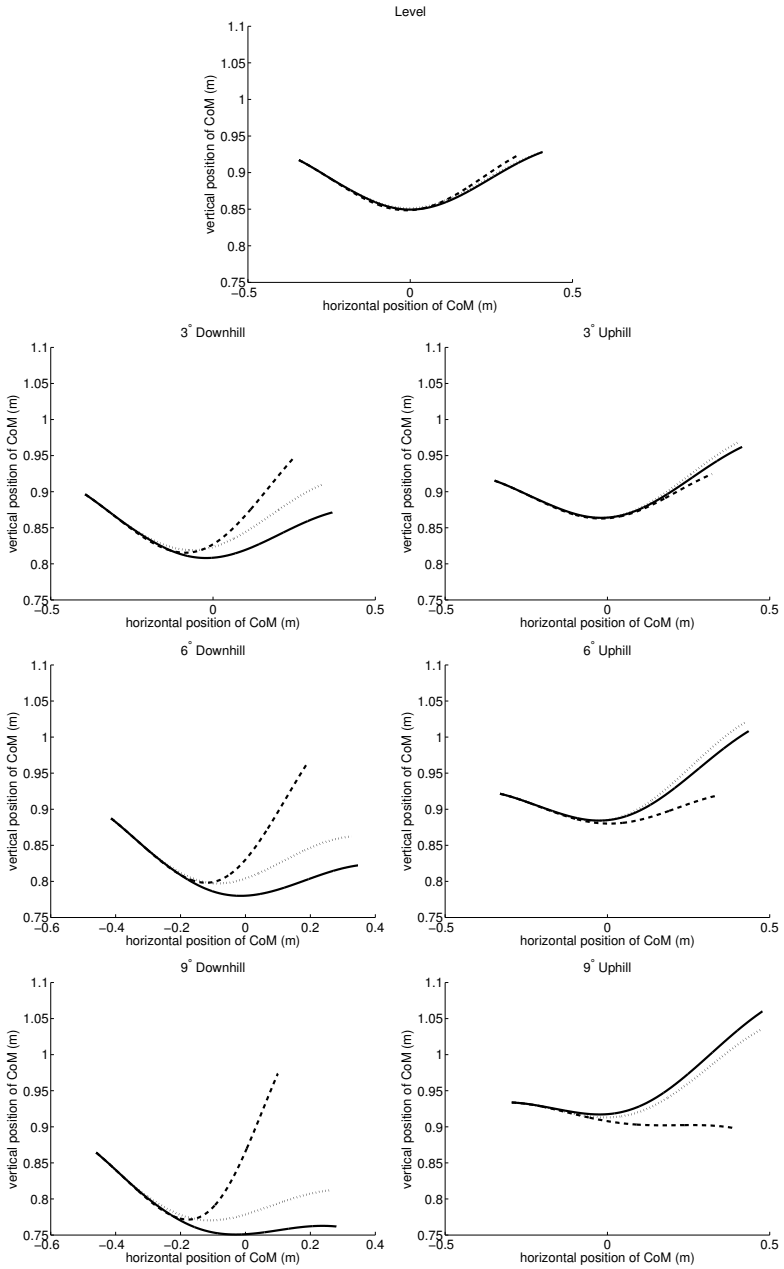

Figure 5.4: Validation data for all slopes.

The validation data shows trajectory of the CoM for the experimental data (solid), the results for the model with no actuation (dashed), and the model with the validation actuation (dotted) for a representative subject for all slopes studied. These trajectories reveal a good overall fit for the uphill. However, the downhill validation data shows a tendency to overestimate the actuation, leading to higher than actual center of mass location.

incline, particularly when looking at the difference between the actuation necessary on hills

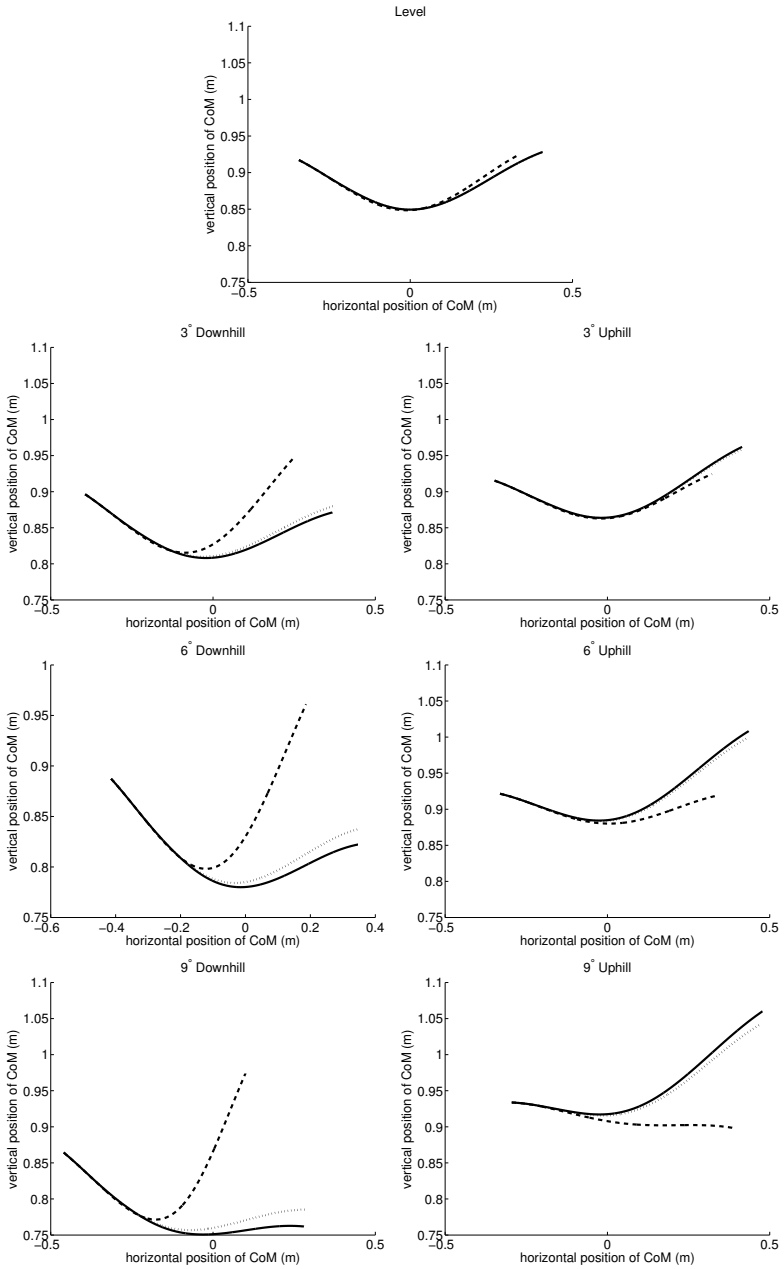

Figure 5.5: Identification data for all slopes.

It is apparent from the data showing the trajectory for the CoM for the experimental data (solid), the results for the model with no actuation (dashed), and the model with the identification actuation (dotted) for the same subject for all slopes that the model is clearly capable of correctly modeling the CoM dynamics on all slopes.

versus on the level. There is very little change in length actuation during the first 20% of

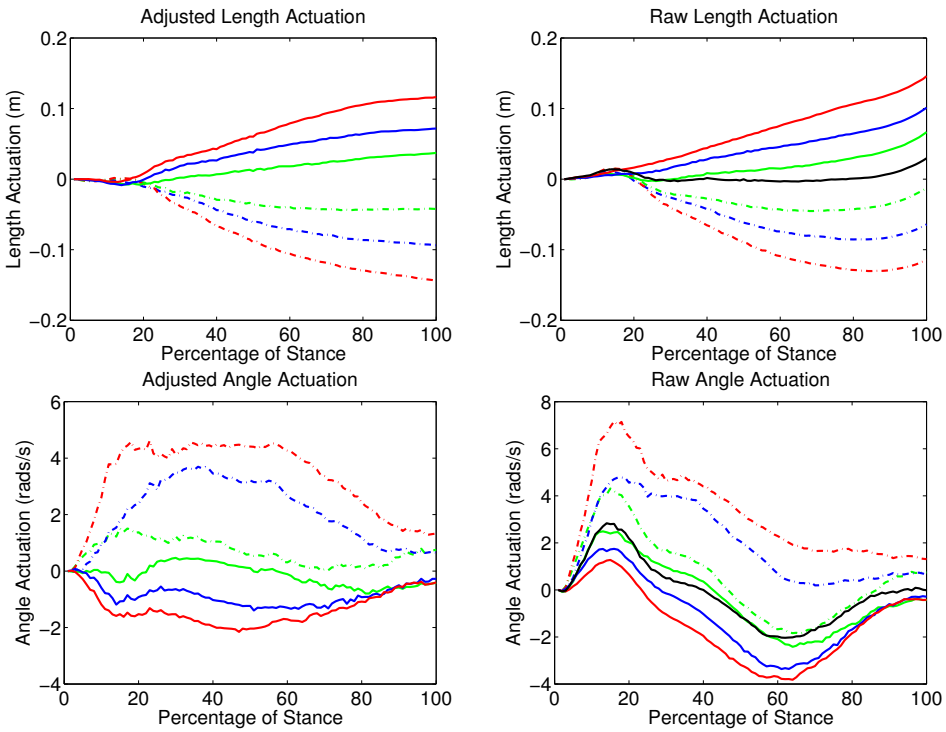

Figure 5.6: Average actuation patterns over stance

The actuation patterns show distinct changes with slope. In all graphs, dash-dotted lines represent downhill, and the solid lines uphill, whereas black is level, green is 3°, blue is 6°, and red is 9°. The top left graph shows the mean difference between the angle actuation on a given uphill or downhill slope and the angle actuation on the level, whereas the top right graph shows the mean raw angle actuation for all slopes, including level. This angle actuation increases with decreasing slope and modestly increases with increasing slope. The bottom left graph shows the mean difference between the length actuation on a on a given uphill or downhill slope and the length actuation on the level, whereas the bottom right graph shows the mean value for the length actuation for all slopes, including level. The length actuation shows the same pattern for all slopes, with only a change in actuation magnitude with slope.

stance, during which elastic energy is stored in all cases. At 20% of stance there is a gradual, approximately linear change in length actuation up until approximately 80% of stance, at which time the change in length tapers off until the end of stance. An increase in incline leads to an increase in the slope of the actuation between 20 and 80%. Similarly, a decrease in slope leads to a decrease in the rate of length actuation.

The angle actuation shows a different pattern (Figure 5.6). There is only a slight

change in angle actuation between the level and the uphill data. Notably, we find virtually no difference in the actuation necessary for level running and for running at a 3° uphill slope. Though there are differences between level running and 6° and 9° uphill slopes, they are small compared to the differences found between the level and the downhill slopes. We also find a relatively smaller difference between the angular actuation on the level and 3° downhill. However, there is then a large change in the actuation needed to go down 6° and 9°.

The variability of the actuation changes across conditions, but the pattern of variability is different for the length actuation than it is for angle actuation (Figures 5.7 and 5.8). While not apparent in a time series, it can be seen when the mean of the standard deviation is taken across the course of stance phase. The angular actuation has low variability for shallow uphill slopes, but the variability increases as with increasing downhill grade and for the steepest uphill grade. The length actuation variability decreases from steep uphill values to shallow uphill values, but has a high variability for all downhill slopes.

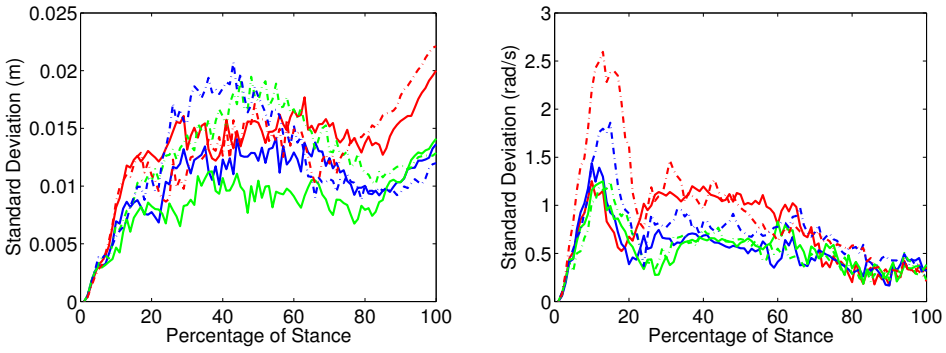

Figure 5.7: Standard deviation patterns across stance

For the length actuation, standard deviations show a similar overall pattern across stance regardless of slope. For the angle actuation, we find large differences in standard deviation during the beginning of stance, likely due to heelstrike, but more modest differences across the rest of stance.

We found a significant decrease in touchdown angle with an increase in slope (Figure 5.9). These differences were greater, however for the downhill angles than the uphill angles.

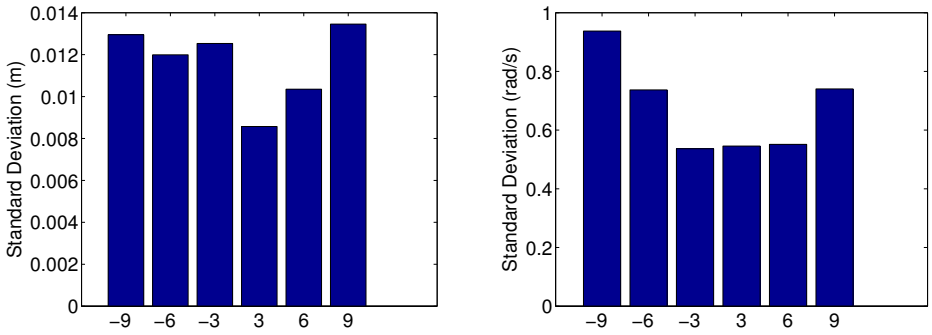

Figure 5.8: Average standard deviation patterns across stance

The mean standard deviation for the entire stance phase shows more variability in actuation downhill than uphill. The one exception is that 9° uphill generally shows just as much variability as do the downhill data.

There was no significant difference between touchdown angle on the level and at 3° uphill, and though differences were found between 3° and 6°, and 6° and 9° degrees downhill, no such differences were found uphill.

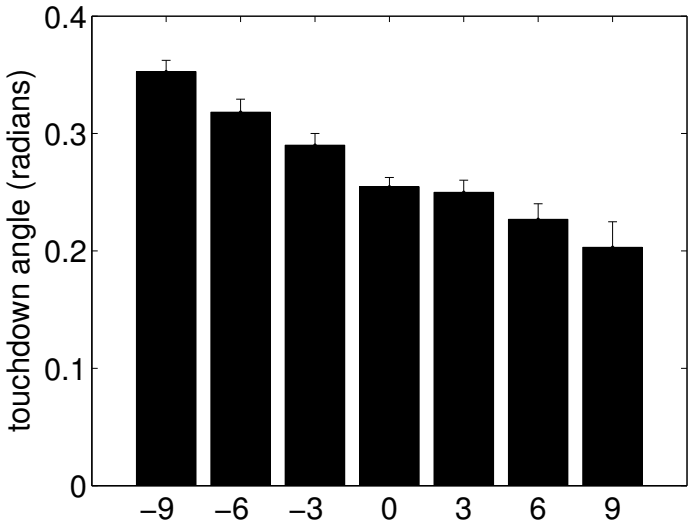

Figure 5.9: Average touchdown angle

Average touchdown angle changes distinctly across slope. Touchdown angle increases distinctly with decreasing slope. Touchdown angle does not change between level running and 3° uphill, but increases with increasing slope for 6° and 9°.

Given these results, we did a sensitivity analysis based on touchdown angle to determine how much these changes might affect actuation across slope. We varied touchdown angle from 40% to 160% of the experimentally identified value for each condition. Touchdown angle affected both the magnitude of the length actuation and that of the angle actuation. An increase in the magnitude of touchdown angle decreased the amount of length actuation necessary to match the experimental trajectory, though the general pattern of increasing length actuation with time for uphill slopes and decreasing length actuation with time for downhill slopes stayed consistent (Figure 5.10). Additionally, touchdown angle affected angle actuation as well, with a decrease in touchdown angle leading to an overall increase in angle actuation, though the general pattern stayed reasonably consistent (Figure 5.11). All of these patterns were consistent across slope.

We also tested the effect of stiffness, by varying it from 40% to 160% of the experimentally identified value for preferred stride frequency on the level. This testing revealed that stiffness significantly affected the pattern of length actuation necessary on a given incline. An increase in stiffness led to a decrease in the actuation necessary to match the experimental trajectory (Figure 5.12). Additionally, while a decrease in stiffness led to a large change in the actuation pattern, an increase led to much smaller changes. However, stiffness had little effect on the angle actuation (Figure 5.13). All of these patterns were consistent across slope

5.4 Discussion

As expected, the model was able to capture the actual changes found between level and uphill and downhill running. The length actuation showed a consistent pattern whether the slope was uphill or downhill, showing a change in magnitude but not pattern with a change in incline or decline. Furthermore, there was little actuation change during the first 20% or the last 20% of stance. The angle actuation showed a sinusoidal pattern for the uphill and 3° downhill, with positive actuation occurring at the very beginning of stance and negative

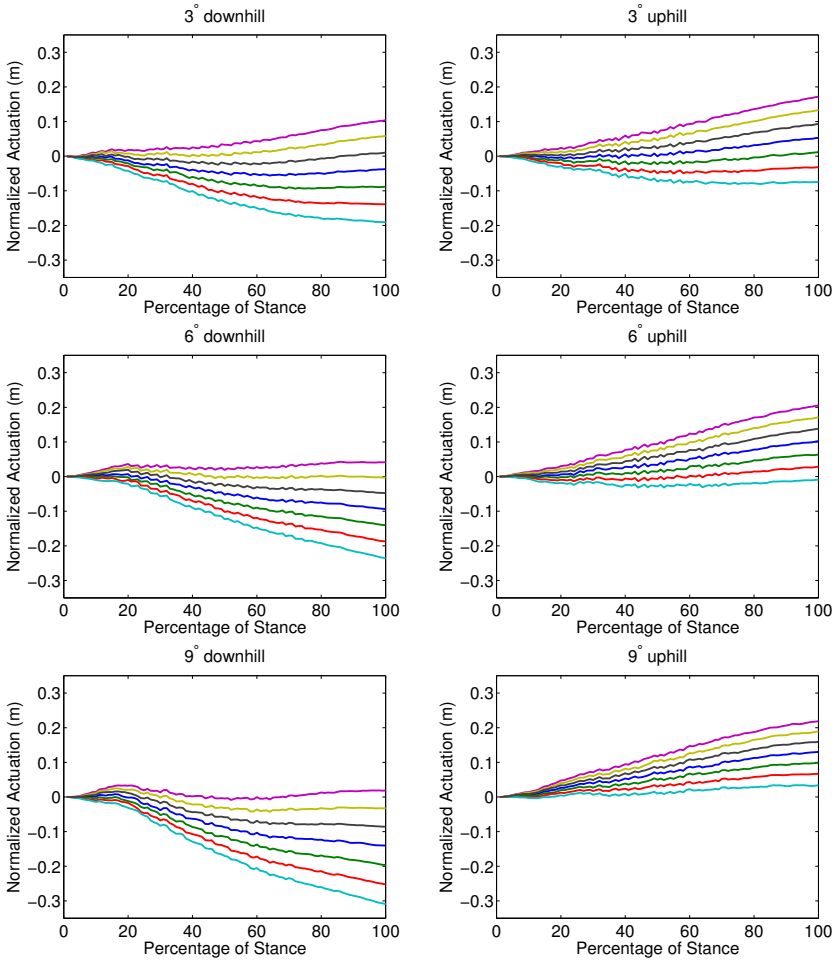

Figure 5.10: Effect of touchdown angle on length actuation for all slopes

Regardless of slope, the mean difference between the length actuation for various touchdown angle values on each non-level slope and experimentally identified touchdown angle affects length actuation similarly. The purple line represents the length actuation for 40% actual stiffness, yellow 60%, grey 80%, blue 100%, green 120%, red, 140% and light blue 160%. Regardless of slope, an increase in touchdown angle leads to an decrease in the need for length actuation.

actuation occurring during the rest of stance. 6° and 9° downhill showed positive actuation all through stance, with the largest values occurring in the first 30% and then tapering off. This is expected because there is generally more muscle activation in the middle of the stance phase, with less occurring at the beginning or the end. Electromyographic (EMG) patterns, which show the electrical activity of the muscle, show greatest activation at mid-

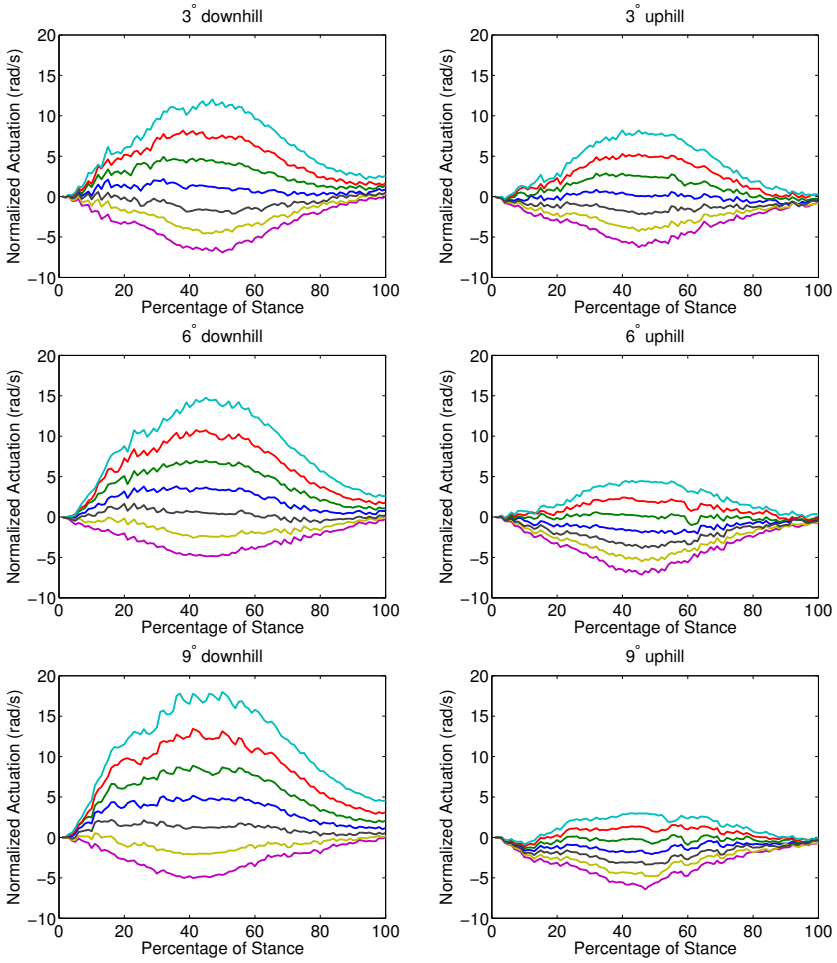

Figure 5.11: Effect of touchdown angle on angle actuation.

Regardless of slope, the mean difference between the length actuation for various touchdown angle values on each non-level slope and experimentally identified touchdown angle affects angle actuation similarly. The purple line represents the angle actuation for 40% actual stiffness, yellow 60%, grey 80%, blue 100%, green 120%, red, 140% and light blue 160%. An increase in touchdown angle leads to an increase in the need for angle actuation.

stance for uphill running, which would correspond to the patterns found at higher stiffness values (Sloniger et al., 1997; Yokozawa et al., 2007; Cappellini et al., 2006; Nilsson et al., 1985). It is possible that the approximately linear change in length that occurs in during the middle of the stance phase corresponds to muscle work that is done during this period of stance.

We expected to find changes in kinematics at shallower slopes that might serve to

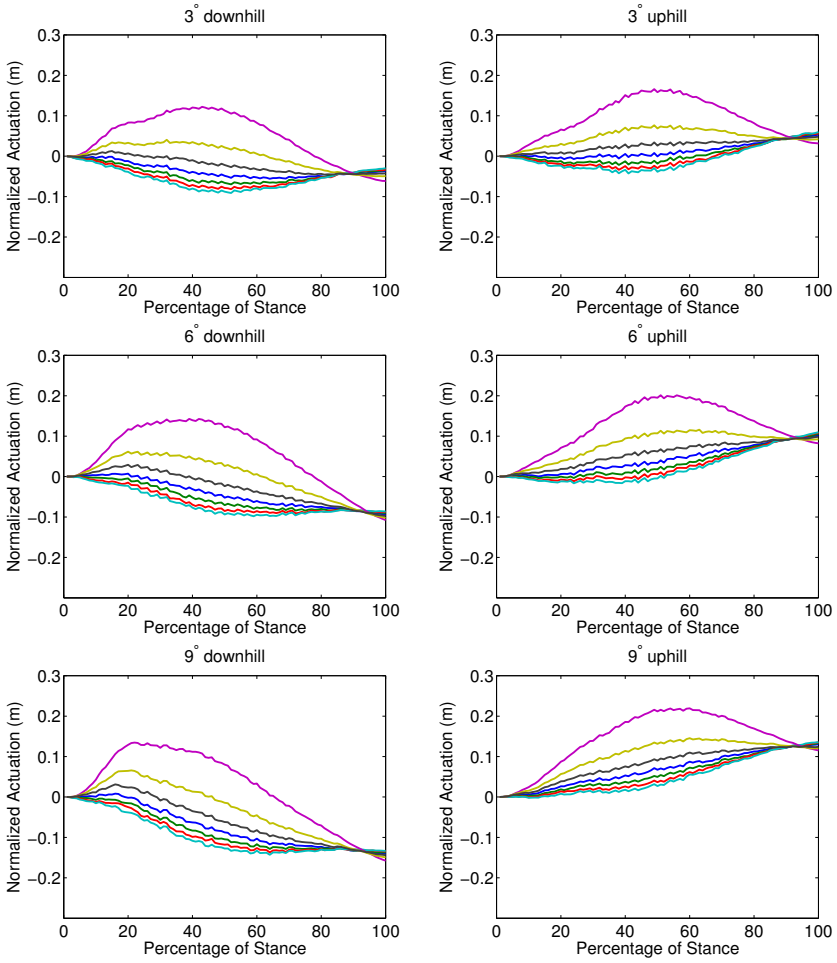

Figure 5.12: Effect of stiffness on length actuation.

The mean difference between the length actuation for various stiffness values on each non-level slope and experimentally identified stiffness on the level shows a relatively consistent pattern regardless of slope . The purple line represents the actuation for 40% actual stiffness, yellow 60%, grey 80%, blue 100%, green 120%, red, 140% and light blue 160%. We find a distinct change in pattern with stiffness. For smaller stiffness values, we find an increase in length actuation followed by a decrease. For greatr stiffness values, we find a decrease in length actuation followed by an increase. We also find smaller changes in length for a given increase in stiffness than for a given decrease in stiffness.

minimize the length actuation necessary. But we did not generally find these patterns. Touchdown angle changed little between the level and 3° uphill, but touchdown angle changed a great deal during downhill running. This may suggest that this relatively simple model lacks the detail to capture the dynamics that might influence such a change.

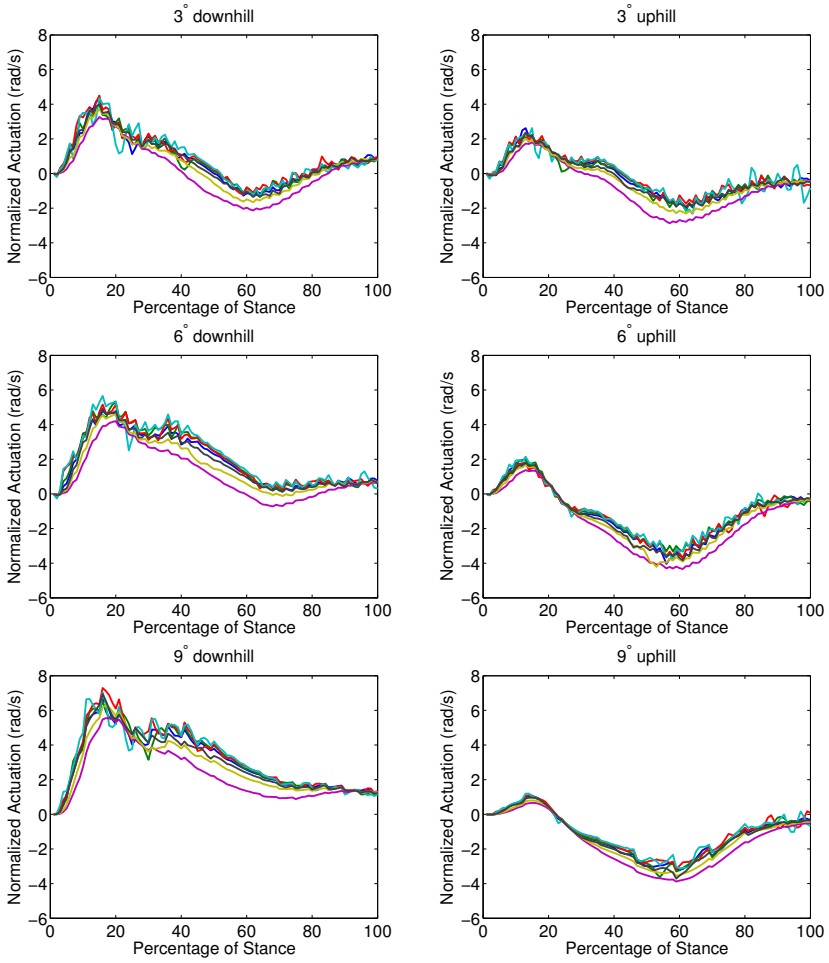

Figure 5.13: Effect of stiffness on angle actuation.

The mean difference between the angle actuation for various stiffness values on each non-level slope and experimentally identified stiffness on the level shows little change regardless of slope. The purple line represents the actuation for 40% actual stiffness, yellow 60%, grey 80%, blue 100%, green 120%, red, 140% and light blue 160%.

Though we did not find the changes we expected to find at shallow slopes, we did get some insight from the sensitivity analysis as to what parameter changes could change actuation patterns. We found that an increase in stiffness led to a decrease in the need for positive length actuation, and a change in its pattern, but little to no change in angle actuation. Typically in humans, increasing slope leads to a modest increase in stride frequency (Minetti, 1994), a kinematic variable that can be represented by an increase in the stiffness

parameter. It is possible that people choose to use faster stride frequencies uphill in order to minimize the positive work that has to be done along the leg.

We also found changes in the actuation patterns with touchdown angle that could influence behavior uphill and downhill. An increase in touchdown angle led to a decrease in the length actuation necessary for a given trajectory, but an increase in angle actuation necessary. The patterns we find of what touchdown angle and actuation patterns people choose to use could represent the different strategies the body uses uphill as opposed to downhill.

Uphill, we found similar angle actuation patterns to what occurred in level running, with only a small change in magnitude with an increase in slope but significant changes in length actuation magnitudes. It is thought that the most expensive part of going from the level to the uphill is the increase in the need to produce positive work via concentric muscle contraction (Margaria, 1976; Minetti et al., 1994). There is also some evidence of bias toward energy generation as opposed to dissipation in non-level running (DeVita et al., 2008). It is possible that the body is maximizing the efficiency of muscle work uphill by producing as much work as possible along the leg while it is close to vertical, with very little occurring in the direction perpendicular to the leg.

We also found small touchdown angles uphill, which minimize the amount of angular actuation that has to occur, but require more length actuation. Hoogkamer and Kram have previously proposed a model for the energy cost of running for a given slope and speed that shows the the cost of work parallel to the treadmill actually decreases with increasing slope (Hoogkamer, 2010). It is possible that the decrease we find in actuation at the ankle corresponds to the decreased metabolic cost of parallel running (the cost incurred for running parallel to the treadmill, which does not include much of the vertical component) whereas the change in length actuation corresponds to the need for positive work. Though it is important to be cautious, because in our case we looked at horizontal as opposed to parallel forces, it possible that the negative actuation we find in uphill running corresponds to this

decrease in metabolic cost.

In downhill running, we find a combination of the same pattern but opposite sign of length actuation that occurs uphill with the addition of significant angle actuation as well as an increase in touchdown angle. During downhill running, to remain at a constant speed, the body needs to dissipate energy. Eccentric action of the muscles, which is used to dissipate energy, is metabolically inexpensive (Abbott et al., 1952), but carries a risk of injury (Armstrong et al., 1983). Therefore, downhill, the focus is likely less on the minimization of energy cost than the prevention of injury to the muscles and joints. There has been experimental evidence of this in walking (Hunter et al., 2010). If we saw a small range of angles swept downhill, as was the case uphill, it would require all of the energy to be dissipated in virtually the same direction, which may overtax or injure some muscles or risk the shock being transmitted to the joints. The increase in touchdown angle and angle actuation, though it may carry a small metabolic penalty, could represent the body's strategy for minimizing injury risk by maximizing the angle swept. We also know that the slope that minimizes energy cost in running occurs somewhere between -6° and -9° (Cavagna et al., 1964; Margaria, 1976; Minetti, 1994). It is possible that the angle actuation allows for energy dissipation to occur across a wider range of angles. If the touchdown angle reaches a limit somewhere around -6° or -9°, this could require a different strategy for steeper declines, one that incurs a larger metabolic cost.

Though this model clearly captures the overall story of what is occurring in uphill and downhill running, it can be improved to reveal more about the underlying dynamics of the system. An improved model could help solve the question of how much of the metabolic cost of level running is due to mechanical work, and how much of that cost is due to isometric force production. Minetti (1994) has argued that mechanical work dominates the cost of running, even on the level. Kram and Taylor, however, have argued that the energy cost of level running is predominantly due to isometric force production, with most of the mechanical work being stored in and then returned from the elastic tendons (Kram and Taylor, 1990).

By adjusting the stiffness values to account for only the amount of elastic energy that is estimated to be stored in the elastic tendons, we could analyze the length and angle actuation necessary in the different slopes. As was found in the sensitivity analysis to stiffness (Figure 5.12), an increase in leg stiffness leads to the need for positive and negative length actuation in both uphill and downhill running. That could represent the need for both the concentric and eccentric contraction that seems to occur in the muscles during uphill and downhill running. By comparing the relative contributions of concentric and eccentric contraction to metabolic cost, it may provide further evidence of how much elastic energy storage occurs across both slope and stride frequency. We have done some preliminary analysis in this area, using stiffnesses 150% and 200% the value of the experimentally identified level stiffness (Figures 5.14 and 5.15).

Furthermore, because this model is relatively simple, it cannot capture what occurs at the different joints throughout the stance phase. However, we collected the necessary kinematic data to do a joint by joint analysis. By looking at the change in joint angle with the torque on the joint on the level, we could calculate the stiffness a spring that represents the work done by each joint on the level. Angle actuation can be added to the model at each joint on the uphill, and solved for using either the methods outlined here or in Chapter 5, similar to the analysis done by (Lee et al., 2008). This would provide further data, not only as to what the magnitude of the actuation is with slope, but which joints modulate the change from level to uphill and level to downhill running.

More accurately understanding how slope affects joint muscle work might allow us to analyze and improve prosthetic leg design. Up until now, lower-leg prosthetic designs have been purely passive, but recently researchers have been exploring the idea of powered prosthetics for walking (Herr and Kornbluh, 2004). One of the most significant losses for faced by amputees is the ability to produce force with muscles. By incorporating actuating elements into the prosthetic devices, engineers could help amputees to be more mobile. Being able to clarify the magnitudes and patterns of force production from a more general

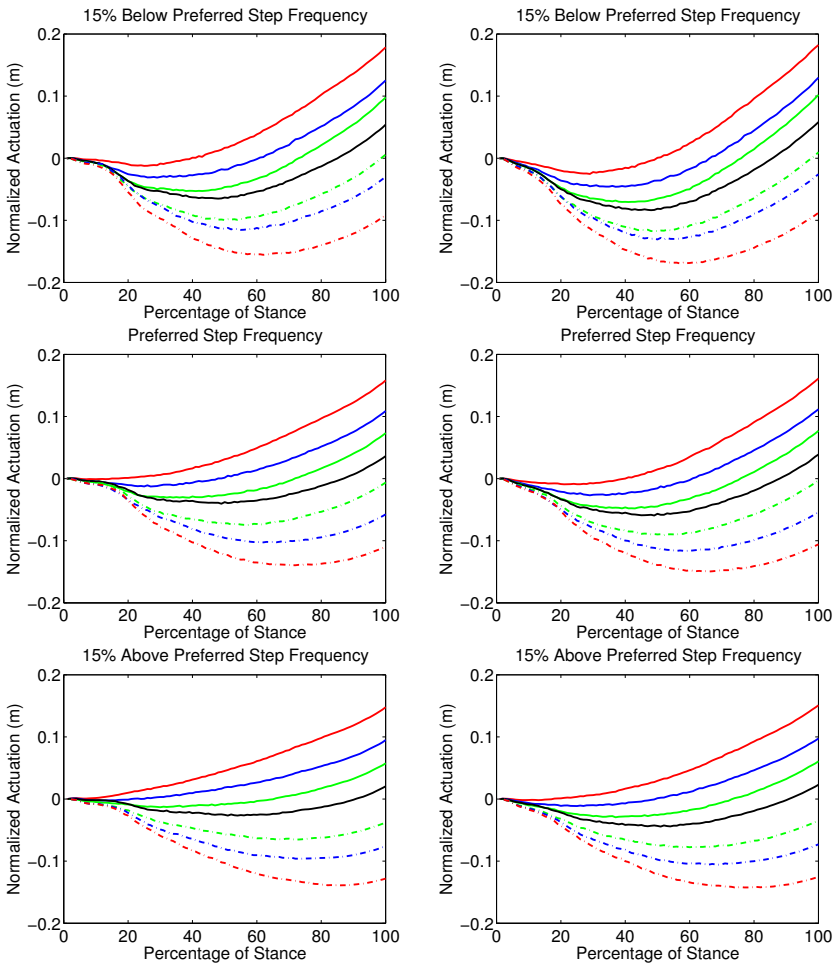

Figure 5.14: Length actuation for modified stiffness values.

With an increased stiffness, we find an increase in the amount of positive length actuation that has to be done on the level, as well as other slopes. In all graphs, dash-dotted lines represent downhill, and the solid lines uphill, whereas black is level, green is 3°, blue is 6°, and red is 9°. There is a greater reduction in actuation for a stiffness of 200% of the experimentally identified value (right hand column) than 150% (left hand column). The amount of actuation that must occur then decreases with increasing stride frequency and increases for decreasing stride frequency. Because these plots depict what occurs for all slopes, not just the change between level and sloped running, we do not normalize based on what occurs on the level.

mechanical standpoint may enable us to effectively construct powered prostheses.

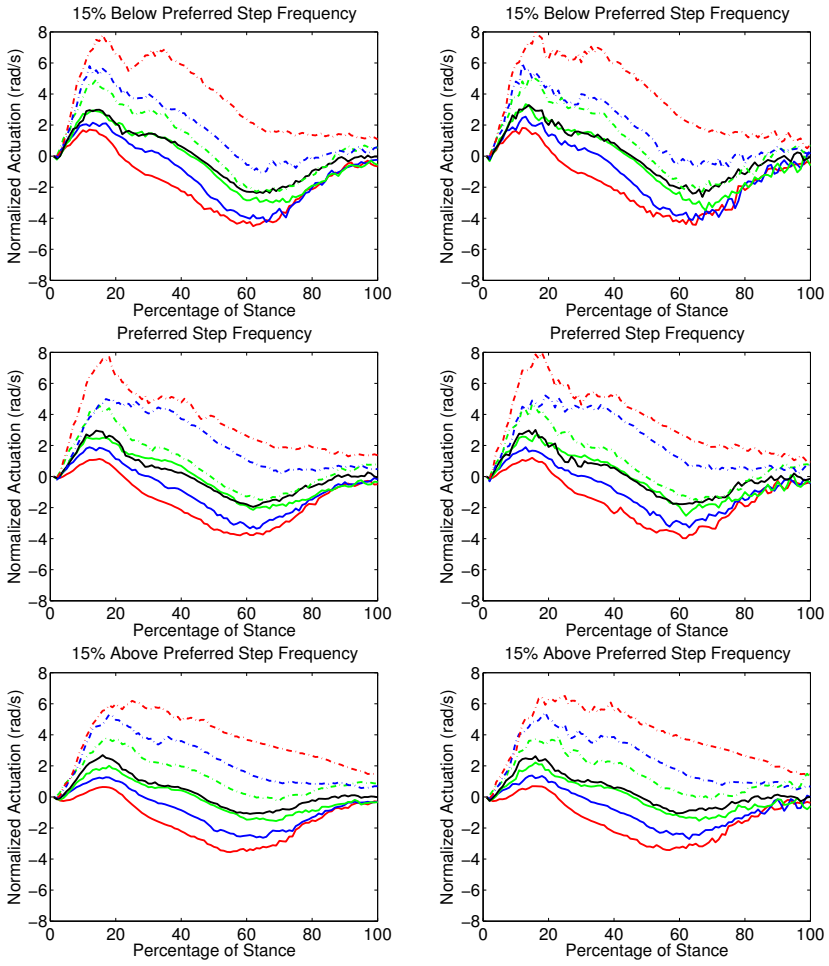

Figure 5.15: Angle actuation for modified stiffness values.

With an increased stiffness, we find little change in the angle actuation that has to be done on the level. The pattern of actuation also changes little with stride frequency, but, as shown earlier, changes significantly with decreasing slope. In all graphs, dash-dotted lines represent downhill, and the solid lines uphill, whereas black is level, green is 3°, blue is 6°, and red is 9°.

Chapter 6

Conclusion

In the previous chapters, I explored a number of different aspects of tuning and control of locomotion, some using and exploring mathematical methods of analyzing experimental data and some exploring mathematical techniques for examining control.

In Chapter 1, we explored the time scales involved in metabolic minimization using step frequency as a correlate for metabolic cost. We found that there were two distinct time scales involved in this process, one fast and one slow. These same two processes were found regardless of the direction perturbation, whether the perturbation was induced by a change in treadmill speed or by imposing and releasing step frequency, and whether or not the initial reaction was an overshoot or an undershoot of the metabolic minimum. We identified some additional dynamics that occurred as a result of a physical perturbation in the form of a fast change in treadmill speed, but these dynamics occurred in addition to those identified for non-physical perturbations, rather than replacing those dynamics. Mathematical modeling revealed that this additional time scale indicated that it could reduce metabolic cost by up to 4.4%, which represents a metabolic savings up as much as 44 Watts. Further research will explore whether or not these time scales also occur in a novel situation.

In Chapter 2, we used previously employed methods utilizing nonlinear dynamics ideas to track stability across walking speed, running speed, and at the approximate transition from walking to running in both modes of locomotion. Our results for this study showed that differences in measure and direction lead to differing results in stability. Additionally,

we found previously employed methods for using nonlinear dynamics to analyze gait are very sensitive to changes in parameters and to data processing techniques. Because past results have examined different variables and used a wide variety of analysis techniques, it is unlikely that it would be possible to compare measures across studies. Furthermore, the lack of a linear region in the average log of the divergence between nearest neighbors indicates that this analysis may not be appropriate to use at all. We plan to do further research using more sophisticated techniques to determine whether or not these results are appropriate, and if so what kind of parameters lead to accurate results. If these measures prove to be repeatable and interpretable, we will compare them to other recently introduced measures of stability (Hof et al., 2005).

In Chapter 3, we developed an alternate method for solving problems using Discrete Mechanics of Optimal Control (DMOC). We used the parallel structure of the underlying constrained optimization problem found in order to reformulate the problem as an unconstrained optimization problem. Using Newton's method, we were obtain quadratic convergence, whereas gradient descent for this problem can only lead to linear convergence and sequential quadratic programming (SQP) only leads to super-linear convergence. Similar results were found whether the system had an unconstrained endpoint, an endpoint constrained by a terminal cost, or an endpoint constrained via projection. We also examined how this method performed for a relatively simple system involving impacts. For more complex systems, such as those involving impacts, Newton's method had insufficiently large basins of attraction. Modifying the algorithm to use trust-region methods when the Hessian was not positive definite and Newton when it was allowed for convergence. Projection shows similar convergence characteristics for systems with impacts, though SQP stalls for some initial conditions. Further research will eliminate some simplifications used in the impact systems, and explore whether or not this method continues to be advantageous for more complex systems, such as those used to model locomotion.

In Chapter 4, we introduce a fully actuated model to analyze the changes involved

in leg mechanics in order to run up or downhill. Starting with the traditional spring-mass model, we added controllers that adjust leg length and leg angle through the stance phase of running. This model revealed relatively consistent patterns in terms of leg length change with uphill and downhill running. There is little actuation required during the first 20% of stance, an approximately linear change for the next 60% of stance, and then a drastic reduction for the last 20% of stance. The magnitude and sign of this change varies with slope, but the pattern is consistent. However, the angle actuation required uphill changed very little with slope, whereas the angle actuation required to go downhill changed significantly with slope, This difference may represent the different strategies involved in uphill and downhill running. Results for uphill running show that most actuation occurs along the leg, and furthermore, that leg angle changes little during stance, indicating that the body may be maximizing the efficiency at which it can produce the positive work necessary for uphill running. Results for downhill running show that leg angle changes significantly throughout stance, with actuation being required to control this change. This could represent the body's strategy of dissipating force at a range of angles, which allows the distribution of this force dissipation over a range of leg muscles. Future analysis will focus on what occurs at each of the joints of the lower limb, determining how much energy is produced and dissipated at each joint.

Each of these studies contributes to our knowledge of how the body functions in a changing environment. Some studies examine the changes that occur when speed, step frequency, or slope changes. Other studies test techniques currently in use to compare stability under different conditions, or offer faster methods for determining optimal control. Future research will continue to explore how we can both experimentally identify and mathematically model how the body adapts to changing circumstances.

Bibliography

Abbott, B., Bigland, B., Ritchie, J., 1952. The physiological cost of negative work. The Journal of Physiology 117 (3), 380.

Alexander, R., 1989. Optimization and gaits in the locomotion of vertebrates. Physiological Reviews 69 (4), 1199.

Altendorfer, R., Koditschek, D., Holmes, P., 2004. Stability analysis of a clock-driven rigid-body SLIP model for RHex. The International Journal of Robotics Research 23 (10-11), 1001.

Altendorfer, R., Moore, N., Komsuoglu, H., Buehler, M., Brown, H., McMordie, D., Saranli, U., Full, R., Koditschek, D., 2001. Rhex: A biologically inspired hexapod runner. Autonomous Robots 11 (3), 207–213.

Arampatzis, A., Brüggemann, G., Metzler, V., 1999. The effect of speed on leg stiffness and joint kinetics in human running. Journal of biomechanics 32 (12), 1349–1353.

Armstrong, R., Ogilvie, R., Schwane, J., 1983. Eccentric exercise-induced injury to rat skeletal muscle. Journal of Applied Physiology 54 (1), 80.

Arnold, A., Asakawa, D., Delp, S., 2000. Do the hamstrings and adductors contribute to excessive internal rotation of the hip in persons with cerebral palsy? Gait and Posture 11 (3), 181–190.

Arnold, A., Thelen, D., Schwartz, M., Anderson, F., Delp, S., 2007. Muscular Coordination of Knee Motion During the Terminal Swing Phase of Normal Gait. Journal of biomechanics 40 (15), 3314.

Bellville, J., Whipp, B., Kaufman, R., Swanson, G., Aqleh, K., Wiberg, D., 1979. Central and peripheral chemoreflex loop gain in normal and carotid body-resected subjects. Journal of Applied Physiology 46 (4), 843.

Berger, W., Dietz, V., Quintern, J., 1984. Corrective reactions to stumbling in man: neuronal co-ordination of bilateral leg muscle activity during gait. The Journal of Physiology 357 (1), 109.

Blickhan, R., 1989. The spring-mass model for running and hopping. Journal of Biomechanics 22 (11-12), 1217.

Blickhan, R., Seyfarth, A., Geyer, H., Grimmer, S., Wagner, H., Günther, M., 2007. Intelligence by mechanics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365 (1850), 199–220.

Bramble, D., Lieberman, D., 2004. Endurance running and the evolution of Homo. Nature 432 (7015), 345–352.

Capelli, C., Rosa, G., Butti, F., Ferretti, G., Veicsteinas, A., di Prampero, P., 1993. Energy cost and efficiency of riding aerodynamic bicycles. European journal of applied physiology and occupational physiology 67 (2), 144–149.

Cappellini, G., Ivanenko, Y., Poppele, R., Lacquaniti, F., 2006. Motor patterns in human walking and running. Journal of neurophysiology 95 (6), 3426.

Carrier, D., Kapoor, A., Kimura, T., Nickels, M., Scott, E., So, J., Trinkaus, E., 1984. The Energetic Paradox of Human Running and Hominid Evolution [and Comments and Reply]. Current Anthropology 25 (4), 483–495.

Cavagna, G., 1975. Force platforms as ergometers. Journal of Applied Physiology 39 (1), 174.

Cavagna, G., Heglund, N., Taylor, C., 1977. Mechanical work in terrestrial locomotion: two basic mechanisms for minimizing energy expenditure. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 233 (5), 243–261.

Cavagna, G., Saibene, F., Margaria, R., 1963. External work in walking. Journal of Applied Physiology 18 (1), 1–9.

Cavagna, G., Saibene, F., Margaria, R., 1964. Mechanical work in running. Journal of Applied Physiology 19 (2), 249.

Cavanagh, P., Williams, K., 1982. The effect of stride length variation on oxygen uptake during distance running. Medicine & Science in Sports & Exercise 14 (1), 30.

Chavarren, J., Calbet, J., 1999. Cycling efficiency and pedalling frequency in road cyclists. European journal of applied physiology and occupational physiology 80 (6), 555–563.

Collins, S., Ruina, A., 2006. A bipedal walking robot with efficient and human-like gait. In: Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on. IEEE, pp. 1983–1988.

Collins, S., Ruina, A., Tedrake, R., Wisse, M., 2005. Efficient bipedal robots based on passive-dynamic walkers. Science 307 (5712), 1082.

Collins, S., Wisse, M., Ruina, A., 2001. A three-dimensional passive-dynamic walking robot with two legs and knees. The International Journal of Robotics Research 20 (7), 607.

Conn, A., Gould, N., Toint, P., 2000. Trust-region methods. Society for Industrial Mathematics.

De Smet, K., Segers, V., Lenoir, M., De Clercq, D., 2009. Spatiotemporal characteristics of spontaneous overground walk-to-run transition. Gait & posture 29 (1), 54–58.

Delp, S., Anderson, F., Arnold, A., Loan, P., Habib, A., John, C., Guendelman, E., Thelen, D., 2007. OpenSim: open-source software to create and analyze dynamic simulations of movement. Biomedical Engineering, IEEE Transactions on 54 (11), 1940–1950.

Delp, S., Loan, J., 2002. A computational framework for simulating and analyzing human and animal movement. Computing in Science & Engineering 2 (5), 46–55.

DeVita, P., Janshen, L., Rider, P., Solnik, S., Hortobágyi, T., 2008. Muscle work is biased toward energy generation over dissipation in non-level running. Journal of biomechanics 41 (16), 3354–3359.

Di Prampero, P., Cortili, G., Mognoni, P., Saibene, F., 1979. Equation of motion of a cyclist. Journal of Applied Physiology 47 (1), 201.

Dietz, V., Quintern, J., Sillem, M., 1987. Stumbling reactions in man: significance of proprioceptive and pre-programmed mechanisms. The Journal of Physiology 386 (1), 149.

Dingwell, J., Cusumano, J., 2000. Nonlinear time series analysis of normal and pathological human walking. Chaos: An Interdisciplinary Journal of Nonlinear Science 10, 848.

Dingwell, J., Cusumano, J., Cavanagh, P., Sternad, D., 2001. Local dynamic stability versus kinematic variability of continuous overground and treadmill walking. Journal of Biomechanical Engineering 123, 27.

Dingwell, J., Cusumano, J., Sternad, D., Cavanagh, P., 2000. Slower speeds in patients with diabetic neuropathy lead to improved local dynamic stability of continuous overground walking. Journal of Biomechanics 33 (10), 1269–1277.

Dingwell, J., Kang, H., 2007. Differences between local and orbital dynamic stability during human walking. Journal of biomechanical engineering 129, 586.

Farley, C., Gonzalez, O., 1996. Leg stiffness and stride frequency in human running. Journal of Biomechanics 29 (2), 181–186.

Farley, C., Houdijk, H., Van Strien, C., Louie, M., 1998. Mechanism of leg stiffness adjustment for hopping on surfaces of different stiffnesses.

Farley, C., Morgenroth, D., 1999. Leg stiffness primarily depends on ankle stiffness during human hopping. Journal of Biomechanics 32 (3), 267–273.

Farrell, P., Wilmore, J., Coyle, E., J.E., B., Costill, D., 1993. Plasma lactate accumulation and distance running performance. Medicine & Science in Sports & Exercise 25 (10), 1091.

Fatemian, M., Nieuwenhuijs, D., Teppema, L., Meinesz, S., Mey, A., Dahan, A., Robbins, P., 2003. The respiratory response to carbon dioxide in humans with unilateral and bilateral resections of the carotid bodies. The Journal of Physiology 549 (3), 965–973.

Ferris, D., Liang, K., Farley, C., 1999. Runners adjust leg stiffness for their first step on a new running surface. Journal of Biomechanics 32 (8), 787–794.

Ferris, D., Louie, M., Farley, C., 1998. Running in the real world: adjusting leg stiffness for different surfaces. Proceedings: Biological Sciences, 989–994.

Fox, M., Delp, S., 2010. Contributions of muscles and passive dynamics to swing initiation over a range of walking speeds. Journal of Biomechanics 43 (8), 1450–1455.

Fraser, A., Swinney, H., 1986. Independent coordinates for strange attractors from mutual information. Physical Review A 33 (2), 1134–1140.

Garcia, M., Chatterjee, A., Ruina, A., Coleman, M., 1998. The simplest walking model: Stability, complexity, and scaling. J Biomech Eeng Trans ASME 120 (2), 281–288.

Geyer, H., 2003. Positive force feedback in bouncing gaits? Proceedings of the Royal Society B: Biological Sciences 270 (1529), 2173–2183.

Geyer, H., Seyfarth, A., Blickhan, R., 2005. Spring-mass running: simple approximate solution and application to gait stability. Journal of Theoretical Biology 232 (3), 315–328.

Geyer, H., Seyfarth, A., Blickhan, R., 2006. Compliant leg behaviour explains basic dynamics of walking and running. Proceedings of the Royal Society B: Biological Sciences 273 (1603), 2861–2867.

Ghigliazza, R., Altendorfer, R., Holmes, P., Koditschek, D., 2005. A simply stabilized running model. SIAM review 47 (3), 519–549.

Gottschall, J., Kram, R., 2005. Ground reaction forces during downhill and uphill running. Journal of biomechanics 38 (3), 445–452.

Günther, M., Blickhan, R., 2002. Joint stiffness of the ankle and the knee in running. Journal of biomechanics 35 (11), 1459–1474.

Gutmann, A., Jacobi, B., Butcher, M., Bertram, J., 2006. Constrained optimization in human running. Journal of Experimental Biology 209 (4), 622.

He, J., Kram, R., McMahon, T., 1991. Mechanics of running under simulated low gravity. Journal of Applied Physiology 71 (3), 863.

Hegger, R., Kantz, H., Schreiber, T., 1999. Practical implementation of nonlinear time series methods: The [small-caps TISEAN] package. Chaos: An Interdisciplinary Journal of Nonlinear Science 9, 413.

Herr, H., Kornbluh, R., 2004. New horizons for orthotic and prosthetic technology: artificial muscle for ambulation. Smart Structures and Materials: Electroactive Polymer Actuators and Devices.

Hicks, J., Arnold, A., Anderson, F., Schwartz, M., Delp, S., 2007. The effect of excessive tibial torsion on the capacity of muscles to extend the hip and knee during single-limb stance. Gait & posture 26 (4), 546–552.

Higginson, J., Zajac, F., Neptune, R., Kautz, S., Delp, S., 2006. Muscle contributions to support during gait in an individual with post-stroke hemiparesis. Journal of biomechanics 39 (10), 1769–1777.

Hof, A., 1998. In vivo measurement of the series elasticity release curve of human triceps surae muscle. Journal of biomechanics 31 (9), 793–800.

Hof, A., Gazendam, M., Sinke, W., 2005. The condition for dynamic stability. Journal of biomechanics 38 (1), 1–8.

Hof, A., Geelen, B., Van den Berg, J., 1983. Calf muscle moment, work and efficiency in level walking; role of series elasticity. Journal of biomechanics 16 (7), 523.

Hogberg, P., 1952. How do stride length and stride frequency influence the energy-output during running? European Journal of Applied Physiology and Occupational Physiology 14 (6), 437–441.

Holmes, P., Koditschek, D., Guckenheimer, J., 2006. The dynamics of legged locomotion: Models, analyses, and challenges. Dynamics 48 (2), 207–304.

Holt, K., Hamill, J., Andres, R., 1991. Predicting the minimal energy costs of human walking. Medicine & Science in Sports & Exercise 23 (4), 491.

Holt, K., Jeng, S., Ratcliffe, R., Hamill, J., 1995. Energetic cost and stability during human walking at the preferred stride frequency. Journal of Motor Behavior 27 (2), 164–178.

Hoogkamer, W., 2010. The metabolic cost of uphill running: a new approach. Master's thesis, Vrije University.

Hreljac, A., 1993. Preferred and energetically optimal gait transition speeds in human locomotion. Medicine & Science in Sports & Exercise 25 (10), 1158.

Hunter, L., Hendrix, E., Dean, J., 2010. The cost of walking downhill: Is the preferred gait energetically optimal? Journal of biomechanics.

Hurmuzlu, Y., Basdogan, C., 1994. On the measurement of dynamic stability of human locomotion. Journal of biomechanical engineering 116, 30.

Hurmuzlu, Y., Basdogan, C., Stoianovici, D., 1995. Kinematics and dynamic stability of the locomotion of polio patients. ASME-Publications-BED 31, 37–38.

Iida, F., Minekawa, Y., Rummel, J., Seyfarth, A., 2007. Toward a human-like biped robot with compliant legs. Robotics and Autonomous Systems.

Iida, F., Rummel, J., Seyfarth, A., 2008. Bipedal walking and running with spring-like biarticular muscles. Journal of Biomechanics 41 (3), 656–667.

Ivanenko, Y., Grasso, R., Macellari, V., Lacquaniti, F., 2002. Control of foot trajectory in human locomotion: role of ground contact forces in simulated reduced gravity. Journal of neurophysiology 87 (6), 3070–3089.

Jordan, K., Challis, J., Cusumano, J., Newell, K., 2008. Stability and the time-dependent structure of gait variability in walking and running. Human Movement Science.

Junge, O., Marsden, J., Ober-Blőbaum, S., 2005. Discrete mechanics and optimal control. In: Proceedings of the 16th IFAC World Congress, Prague, Czech Republic.

Kandel, E., Schwarz, J., Jessell, T., 2000. Principles of neural science. McGraw Hill New York.

Kaneko, M., Matsumoto, M., Ito, A., Fuchimoto, T., 1987. Optimum step frequency in constant speed running. Biomechanics XB, 803–807.

Kaufman, M., Hayes, S., 2002. The exercise pressor reflex. Clinical Autonomic Research 12 (6), 429–439.

Kaufman, M., Longhurst, J., Rybicki, K., Wallach, J., Mitchell, J., 1983. Effects of static muscular contraction on impulse activity of groups III and IV afferents in cats. Journal of Applied Physiology 55 (1), 105.

Kelley, C., 1995. Iterative methods for linear and nonlinear equations. Society for Industrial Mathematics.

Kennel, M., Brown, R., Abarbanel, H., 1992. Determining embedding dimension for phase space reconstruction using the method of false nearest neighbors. Physica D.

Kerdok, A., Biewener, A., McMahon, T., Weyand, P., Herr, H., 2002. Energetics and mechanics of human running on surfaces of different stiffnesses. Journal of Applied Physiology 92 (2), 469–478.

Knuesel, H., Geyer, H., Seyfarth, A., 2005. Influence of swing leg movement on running stability. Human movement science 24 (4), 532–543.

Koditschek, D., Full, R., Buehler, M., 2004. Mechanical aspects of legged locomotion control. Arthropod Structure and Development 33 (3), 251–272.

Kram, R., Taylor, C., 1990. Energetics of running: a new perspective. Nature 346 (6281), 265–267.

Kuo, A., 1999. Stabilization of lateral motion in passive dynamic walking. The International Journal of Robotics Research 18 (9), 917.

Kuo, A., 2001. A simple model of bipedal walking predicts the preferred speed–step length relationship. Journal of biomechanical engineering 123, 264.

Kuo, A., 2002. Energetics of actively powered locomotion using the simplest walking model. Journal of Biomechanical Engineering 124, 113.

Lee, D., McGuigan, M., Yoo, E., Biewener, A., 2008. Compliance, actuation, and work characteristics of the goat foreleg and hindleg during level, uphill, and downhill running. Journal of Applied Physiology 104 (1), 130.

Leyendecker, S., Ober-Blőbaum, S., Marsden, J., Ortiz, M., 2007. Discrete mechanics and optimal control for constrained multibody dynamics. In: Proceedings of the 6th International Conference on Multibody Systems, Nonlinear Dynamics, and Control, ASME International Design Engineering Technical Conferences, Las Vegas, Nevada. pp. 4–7.

Lichtwark, G., Bougoulias, K., Wilson, A., 2007. Muscle fascicle and series elastic element length changes along the length of the human gastrocnemius during walking and running. Journal of biomechanics 40 (1), 157–164.

Liu, M., Anderson, F., Schwartz, M., Delp, S., 2008. Muscle contributions to support and progression over a range of walking speeds. Journal of biomechanics 41 (15), 3243–3252.

Ljung, L., 1987. System Identification: Theory for the User Prentice-Hall. Englewood Cliffs, NJ. 519pp.

Luksch, T., Berns, K., Mombaur, K., Schultz, G., 2007. Using optimization techniques for the design and control of fast bipeds. In: 10th International Conference on Climbing and Walking Robots (CLAWAR), Singapore. Citeseer.

Margaria, R., 1976. Biomechanics and energetics of muscular exercise. Clarendon Press Oxford.

Marsden, J., West, M., 2003. Discrete mechanics and variational integrators. Acta Numerica 10, 357–514.

McGeer, T., 1990a. Passive bipedal running. Proceedings of the Royal Society of London. Series B, Biological Sciences (1934-1990) 240 (1297), 107–134.

McGeer, T., 1990b. Passive dynamic walking. The International Journal of Robotics Research 9 (2), 62.

McGowan, C., Kram, R., Neptune, R., 2009. Modulation of leg muscle function in response to altered demand for body support and forward propulsion during walking. Journal of biomechanics 42 (7), 850–856.

McMahon, T., Cheng, G., 1990. The mechanics of running: how does stiffness couple with speed? Journal of Biomechanics 23, 65.

McMahon, T., Valiant, G., Frederick, E., 1987. Groucho running. Journal of Applied Physiology 62 (6), 2326.

Mercier, J., Gallais, D., Durand, M., Goudal, C., Micallef, J., Préfaut, C., 1994. Energy expenditure and cardiorespiratory responses at the transition between walking and running. European journal of applied physiology and occupational physiology 69 (6), 525–529.

Minetti, A., 1994. Mechanical determinants of the minimum energy cost of gradient running in humans. Journal of Experimental Biology 195 (1), 211–225.

Minetti, A., Ardigo, L., Saibene, F., 1994. The transition between walking and running in humans: metabolic and mechanical aspects at different gradients. Acta Physiologica Scandinavica 150 (3), 315–323.

Minetti, A., Capelli, C., Zamparo, P., di Prampero, P., Saibene, F., 1995. Effects of stride frequency on mechanical power and energy expenditure of walking. Medicine & Science in Sports & Exercise 27 (8), 1194.

Mombaur, K., 2007. Human-like Running Can Be Open-Loop Stable. Autonome Mobile Systeme 2007, 282–286.

Mombaur, K., 2009. Using optimization to create self-stable human-like running. Robotica 27 (03), 321–330.

Mombaur, K., Truong, A., Laumond, J., 2010. From human to humanoid locomotionan inverse optimal control approach. Autonomous Robots 28 (3), 369–383.

Neptune, R., Clark, D., Kautz, S., 2009. Modular control of human walking: a simulation study. Journal of biomechanics 42 (9), 1282.

Neptune, R., Sasaki, K., Kautz, S., 2008. The effect of walking speed on muscle function and mechanical energetics. Gait & posture 28 (1), 135.

Nilsson, J., Thorstensson, A., Halbertsma, J., 1985. Changes in leg movements and muscle activity with speed of locomotion and mode of progression in humans. Acta Physiologica Scandinavica 123 (4), 457–475.

Nocedal, J., Wright, S., 1999. Numerical optimization. Springer.

Ober-Bloebaum, S., Junge, O., Marsden, J., 2008. Discrete Mechanics and Optimal Control: an Analysis. Arxiv preprint arXiv:0810.1386.

Pandy, M., Andriacchi, T., 2010. Muscle and Joint Function in Human Locomotion. Annual Review of Biomedical Engineering 12, 401–433.

Pearson, K., 2004. Generating the walking gait: role of sensory feedback. Progress in Brain Research 143, 123–129.

Pekarek, D., Ames, A., Marsden, J., 2007. Discrete mechanics and optimal control applied to the compass gait biped. In: IEEE Conference on Decision and Control and European Control Conference, New Orleans, LA, USA.

Pekarek, D., Marsden, J., 2008. Variational Collision Integrators and Optimal Control. In: Proceedings of the 18th International Symposium on Mathematical Theory of Networks and Systems, Blacksburg, VA.

Roberts, T., Marsh, R., Weyand, P., Taylor, C., 1997. Muscular force in running turkeys: the economy of minimizing work. Science 275 (5303), 1113.

Rosenstein, M., Collins, J., De Luca, C., et al., 1993. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D 65 (1-2), 117–134.

Rummel, J., Iida, F., Seyfarth, A., 2006. One-legged locomotion with a compliant passive joint. In: Proceedings of the International Conference on Intelligent Autonomous Systems Tokyo, Japan. pp. 566–573.

Rummel, J., Iida, F., Smith, J., Seyfarth, A., 2008. Enlarging regions of stable running with segmented legs. In: IEEE International Conference on Robotics and Automation, 2008. ICRA 2008. pp. 367–372.

Rummel, J., Seyfarth, A., 2008. Stable Running with Segmented Legs. The International Journal of Robotics Research 27 (8), 919.

Sasaki, K., Neptune, R., 2006a. Differences in muscle function during walking and running at the same speed. Journal of biomechanics 39 (11), 2005–2013.

Sasaki, K., Neptune, R., 2006b. Muscle mechanical work and elastic energy utilization during walking and running near the preferred gait transition speed. Gait and Posture 23 (3), 383–390.

Segers, V., Aerts, P., Lenoir, M., De Clercq, D., 2006. Spatiotemporal characteristics of the walk-to-run and run-to-walk transition when gradually changing speed. Gait & posture 24 (2), 247–254.

Segers, V., Aerts, P., Lenoir, M., De Clerq, D., 2007. Dynamics of the body centre of mass during actual acceleration across transition speed. Journal of Experimental Biology 210 (4), 578.

Seyfarth, A., Geyer, H., Blickhan, R., Lipfert, S., Rummel, J., Minekawa, Y., Iida, F., 2006. Running and walking with compliant legs. Lecture notes in control and information sciences 340, 383.

Seyfarth, A., Geyer, H., Günther, M., Blickhan, R., 2002. A movement criterion for running. Journal of Biomechanics 35 (5), 649–655.

Seyfarth, A., Geyer, H., Herr, H., 2003. Swing-leg retraction: a simple control model for stable running. Journal of Experimental Biology 206 (15), 2547–2555.

Seyfarth, A., Iida, F., Tausch, R., Stelzer, M., Von Stryk, O., Karguth, A., 2009. Towards bipedal jogging as a natural result of optimizing walking speed for passively compliant three-segmented legs. The International Journal of Robotics Research 28 (2), 257.

Sloniger, M., Cureton, K., Prior, B., Evans, E., 1997. Lower extremity muscle activation during horizontal and uphill running. Journal of Applied Physiology 83 (6), 2073.

Snaterse, M., Ton, R., Kuo, A., Donelan, J., 2011. Distinct fast and slow processes contribute to the selection of preferred step frequency during human walking. Journal of Applied Physiology Accepted.

Snyder, K., Farley, C., 2011. Energetically optimal stride frequency in running: the effect of incline and decline. Journal of Experimental Biology Accepted.

Srinivasan, M., Ruina, A., 2005. Computer optimization of a minimal biped model discovers walking and running. Nature 439 (7072), 72–75.

Srinivasan, M., Ruina, A., 2007. Idealized walking and running gaits minimize work. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science 463 (2086), 2429.

Stefanyshyn, D., Nigg, B., 1998. Dynamic angular stiffness of the ankle joint during running and sprinting. Journal of Applied Biomechanics 14, 292–299.

Takens, F., 1981. Detecting strange attractors in turbulence. Dynamical systems and turbulence, Warwick 1980, 366–381.

Umberger, B., Martin, P., 2007. Mechanical power and efficiency of level walking with different stride rates. Journal of Experimental Biology 210 (18), 3255–3265.

Whelan, P., 1996. Control of locomotion in the decerebrate cat. Progress in Neurobiology 49 (5), 481–515.

Wisse, M., Schwab, A., Van Der Helm, F., 2004. Passive dynamic walking model with upper body. Robotica 22 (06), 681–688.

Wisse, M., Schwab, A., van der Linde, R., van der Helm, F., 2005. How to keep from falling forward: elementary swing leg action for passive dynamic walkers. Robotics, IEEE Transactions on 21 (3), 393–401.

Yokozawa, T., Fujii, N., Ae, M., 2007. Muscle activities of the lower limb during level and uphill running. Journal of Biomechanics 40 (15), 3467–3475.

Zarrugh, M., Radcliffe, C., 1978. Predicting metabolic cost of level walking. European Journal of Applied Physiology and Occupational Physiology 38 (3), 215–223.